



UvA-DARE (Digital Academic Repository)

A state space approach to the estimation of multi-factor affine stochastic volatility option pricing models

van der Ploeg, A.P.C.; Boswijk, H.P.; de Jong, F.

Publication date

2003

[Link to publication](#)

Citation for published version (APA):

van der Ploeg, A. P. C., Boswijk, H. P., & de Jong, F. (2003). *A state space approach to the estimation of multi-factor affine stochastic volatility option pricing models*. (UvA-econometrics working paper; No. 13). UvA.

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Discussion Paper: 2003/13

A state space approach to the estimation of multi-factor affine stochastic volatility option pricing models

Antoine P.C. van der Ploeg, H. Peter Boswijk
and Frank de Jong

www.fee.uva.nl/ke/UvA-Econometrics

Department of Quantitative Economics
Faculty of Economics and Econometrics
Universiteit van Amsterdam
Roetersstraat 11
1018 WB AMSTERDAM
The Netherlands

UvA  UNIVERSITEIT VAN AMSTERDAM



A State Space Approach to the Estimation of Multi-Factor Affine Stochastic Volatility Option Pricing Models

Antoine P.C. van der Ploeg^a ¹, H. Peter Boswijk^a, Frank de Jong^b

^a Department of Quantitative Economics, University of Amsterdam,
Roetersstraat 11, 1018 WB, Amsterdam, The Netherlands

^b Department of Finance, University of Amsterdam, The Netherlands

(10 July 2003)

Abstract

We propose a class of stochastic volatility (SV) option pricing models that is more flexible than the more conventional models in different ways. We assume the conditional variance of the stock returns to be driven by an affine function of an arbitrary number of latent factors, which follow mean-reverting Markov diffusions. This set-up, for which we got the inspiration from the literature on the term structure of interest rates, allows us to empirically investigate if volatilities are driven by more than one factor. We derive a call pricing formula for this class. Next, we propose a method to estimate the parameters of such models based on the Kalman filter and smoother, exploiting both the time series and cross-section information inherent in the options and the underlying simultaneously. We argue that this method may be considered an attractive alternative to the efficient method of moments (EMM). We use data on the FTSE100 index to illustrate the method. We provide promising estimation results for a 1-factor model in which the factor follows an Ornstein-Uhlenbeck process. The results indicate that the method seems to work well. Diagnostic checks show evidence of there being more than one factor that drives the volatility, indicate the existence of level-dependent volatility, and provide an incentive to consider realized volatility in future empirical analysis.

JEL classification: C13, C32, G12, G13.

Keywords: Derivative pricing, Stochastic Volatility, Kalman filter, State space models.

1. Introduction

The problem of parameter estimation in stochastic volatility (SV) option pricing models is generally considered to be difficult. Such a model typically consists of two equations to describe the evolution of the underlying asset on which the option is written; one for the stock price and another for its random volatility. See e.g., the models considered by Hull and White (1987), Wiggins (1987), Scott (1987), Stein and Stein (1991), Heston (1993), Ball and Roma (1994) and many others.

The main issues that arise are the following. Both empirically observed stock and option prices contain complementary information about parameter values. As such, the data has both a time series and cross-section dimension. Stock prices contain information about the parameter values of their real-world distribution. Option prices however, incorporate information about the parameters of the stock's risk-neutral distribution as well, which is a necessary input for pricing derivatives. Moreover, as the market is incomplete in the sense that not all derivatives can be perfectly hedged due to stochastic volatility, the change to the risk-neutral measure is not unique. It

¹ Corresponding author. E-mail address: a.p.c.vanderploeg@uva.nl . Tel. +31 20 525 5269.

is determined by the market price of volatility risk which depends on the risk preferences of the agents in the market. In order to obtain unique derivative prices a certain functional form needs to be assumed for this market price of volatility risk. Besides, both stock and option prices contain complementary information about the unobservable stock volatility. Ignoring one of these sources of information probably leads to a loss. Incorporating both sources is likely to generate more efficient parameter estimates. This is important for more reliable pricing and hedging of, e.g., newly issued exotic over-the-counter derivative products by a financial institution. Furthermore, it may generate better volatility forecasts which are relevant in all kinds of risk-management systems like Value-at-Risk.

However, combining both sources for reliable parameter estimation is difficult as these models deal with a latent object, the volatility. As volatility is unobservable it needs to be integrated out both when pricing derivatives and when estimating the parameters by means of optimizing some fitting criterion. And here the main problems come in: computationally very demanding simulations and numerical multiple integrations are required.

As a consequence the literature has mainly concentrated on the information in either option data or stock price data alone. Studies that investigate the informational content of option prices and contrast them with the information in the underlying are summarized in Bates (1996). This informational content often concerns the stock volatility. If one estimates the volatility from a stock price series, then this estimate is based on historical data alone and may not be too relevant for predicting future volatility². In contrast, implied volatilities obtained from option prices are typically considered to be market forecasts of future volatility, and may therefore be more relevant. To infer diffusion parameters from option price data, a technique called *calibration* is commonly employed. See, e.g., Bates (1996), Bakshi, Cao and Chen (1997) and Duffie, Pan and Singleton (2000). This technique exploits the information in a cross-section (i.e., a panel) of option prices on a certain day, and infers the \mathbb{Q} -parameters from these prices by minimizing the sum of squared deviations between the observed market prices and the theoretical prices. This technique disregards the time-series dimension, and is therefore inherently inconsistent with the dynamic principles of the model.

Another part of the literature has focused on the information in stock prices alone for estimating the diffusion parameters³. Together with some assumption on the market price of volatility risk, derivatives are subsequently priced. For example, Scott (1987) and Taylor (1994) use the method of moments (MM). Others employ generalized method of moments (GMM). Although appealing at first sight, a straightforward application of (G)MM is not always possible. The main reason being that for many SV models a sufficient number of moments cannot analytically be derived. This has led researchers like Wiggins (1987), Chesney and Scott (1989), Melino and Turnbull (1990) and notably Duffie and Singleton (1993), to simulate the unknown moments, and then apply GMM. This method is known as simulated method of moments. Shephard (1996) lists a large number of drawbacks of using GMM for estimating SV models. Especially notable is the fact that it does not deliver a volatility forecast, so

² This does not appear to be true for the *realized volatility* models based on high frequency intraday stock return data, which seem to generate accurate forecasts. For a recent application, see Hol and Koopman (2002). However, in this paper we look at *daily* option and stock price data.

³ It is not our aim to provide a complete survey of all estimation methods that have been proposed in the literature for estimating SV models. Overviews can be found in, inter alia, Ghysels et al. (1996), Shephard (1996) and Van der Sluis (1999).

that another type of estimation is needed for that. Furthermore, it is generally known that GMM may have serious finite sample problems. Another branch of estimation techniques based on stock prices only, explicitly recognizes the fact that a latent process drives the volatility. These techniques write the process in terms of an unobserved components- or *state space* model. Examples include Harvey, Ruiz and Shephard (1994), Kim, Shephard and Chib (1998) and Sandmann and Koopman (1998).

Today however, a method coined *efficient method of moments* (EMM) by Gallant and Tauchen (1996), is probably the most popular method for estimating SV models. EMM matches the scores of the likelihood function of an auxiliary model via simulation. Gallant and Tauchen show that the EMM estimator is asymptotically as efficient as the maximum likelihood estimator, if the auxiliary model is a good approximation to the distribution of the data. EMM has been used primarily for estimating diffusions using only stock price data. Nevertheless, Chernov and Ghysels (2000) have extended EMM to estimate and appraise SV option pricing models using the joint distribution of the underlying stock and its options. They are the first to combine both sources of information simultaneously for estimation purposes. However, EMM is not easy to implement. Another major drawback is that it does not readily deliver a volatility forecast; another method called *reprojection* is required for this. Furthermore, although EMM has desirable asymptotic properties, accumulating evidence from empirical studies, which typically deal with finite samples, point out that the finite sample behavior of this estimation method can be very poor. See, e.g., Gallant, Hsieh and Tauchen (1997), Duffee and Stanton (2001) and Chernov and Ghysels (2000).

The main purpose of this paper is twofold. First, we propose a different estimation technique for estimating SV option pricing models, that permits using both sources of information simultaneously, like Chernov and Ghysels (2000) did. Our technique is based on the Kalman filter. It allows to fully exploit the time series and cross-section information in prices of a panel of options and their underlying value jointly. A convenient consequence of our procedure is that the filtered volatility series is a direct by-product of the estimation output. Moreover, the market price of volatility risk can be isolated in this way. Second, before discussing our technique we start by proposing an SV model that is more flexible than generally encountered in the literature. We assume the stock variance to be driven by an affine function of an arbitrary number of latent factors (instead of just one), which are assumed to follow stationary mean-reverting Markov diffusions. We label the model the *multi-factor affine stochastic volatility option pricing model*. We do not allow for the leverage effect (yet) in our analysis⁴. The reason for proposing this multi-factor model is the recent empirical finding that the term structure of stock volatilities seems to be attributed to more than one factor. See Bates (2000), and Duffie, Pan and Singleton (2000). It would be interesting to empirically investigate by how many factors the volatility is actually driven. Although our technique permits this the empirical illustration in this paper is for a 1-factor model.

The inspiration for setting up such an SV model and our estimation technique was found in the literature dealing with the term structure of interest rates. This literature generally assumes the short interest rate to be driven by an affine function of latent factors, which follow stationary mean-reverting Markov diffusions. See Duffie and Kan (1996) and Dai and Singleton (2000) amongst many others. As

⁴ Incorporating the leverage effect will be considered in future research. Now it would complicate matters too much.

analytical bond pricing formulas result in these models, which is very convenient in terms of parameter estimation, we tried to transfer the insights obtained there towards the problem of estimating SV option pricing models, and to benefit from it.

After having stated the general multi-factor affine SV option pricing model, we derive a valuation formula for call options for this model class. This formula turns out to essentially coincide with the Hull and White (1987) formula. It tells us that in order to obtain the call value, we need to compute the conditional expectation (under the risk-neutral measure Q) of the Black-Scholes (1973) pricing function evaluated in the integrated variance over the remaining life of the option.

The problem with this valuation formula is the fact that an explicit analytical expression does not result, except in a few special cases such as the Heston (1993) model⁵. In general, prices can only be obtained by Monte Carlo simulation. We already argued that this poses a major problem for econometric estimation of such models. To avoid reliance on Monte Carlo simulations during estimation, we propose the following procedure. First, we rewrite the Black-Scholes (BS) pricing function in terms of the exponent of the integrated variance, which we see as the new argument of the BS function. Next, we linearize the BS function around a suitably chosen value of this new argument. Given this linearization of the BS function, and reconsidering the fact that the call price is obtained by taking the conditional Q -expectation of this function, then, if we take the Q -expectation of this linearization, we end up with a partly analytical expression for the call price. Namely, from the results of Duffie and Kan (1996), we know that the Q -expectation of the exponent of the integrated variance is of the exponential-affine form. Some further manipulations to be discussed below then ensure that we end up with an equation that is linear in the latent factors. This equation can then serve as part of the measurement equation of a linear state space model that we will use for parameter estimation. In order to investigate if linearization of the BS function provides a reasonable approximation to the true pricing function, we provide some promising results that indicate that this may indeed be the case.

Tackling the call valuation formula in this way to write it into linear state space form is only part of deriving proper discrete-time equations from our continuous-time option pricing model. The stochastic differential equations (SDEs) for the underlying stock price and the volatility driving factors must also be handled to make them suitable for proper state space estimation. This is covered in detail below.

Having collected together the equations that make up the resulting linear state space model, we turn to an empirical implementation. We examine daily data on the UK's FTSE100 index, and the European option contract traded on that index, covering the period 4-1-1993 – 28-12-2001. We provide some promising initial estimation results for the simplest special case, in which there is one factor driving the volatility that follows an Ornstein-Uhlenbeck process. The model is estimated using three types of data: only stock return data, only option data, and both sources of data. The extracted volatility series is compared to a GARCH volatility series and to the observed Black-Scholes implied volatility data. The graphs we supply indicate that the method may work well, although further research has to confirm this. Diagnostic checks show evidence of there being more than one factor that drives the volatility, indicate the existence of level-dependent volatility, and provide an incentive to consider realized volatility in our analysis. Together with an extension towards

⁵ The Heston (1993) model with correlation parameter $\rho = 0$ is a special case of our model.

incorporating the leverage effect, and towards considering a panel of options, this will all be investigated in our future research.

The remaining part of the paper is structured as follows. Section 2 discusses the general set-up of the multi-factor affine SV option pricing model, and shows how call options can be valued in this setting. Section 3 is devoted to the state space framework. The advantages of the state space approach in the current context are highlighted. A brief review of the linear state space framework, the Kalman filter and smoother, and QML estimation is given next. Section 4 is devoted to writing the model in linear state space form. Each equation from the continuous-time model is considered separately. Special attention is devoted to writing the call pricing formula in the desired format by using a linear approximation. Some evidence on the quality of this linearization is discussed thereafter. Section 5 discusses the data and estimation results for the 1-factor Ornstein-Uhlenbeck special case, including diagnostic checks. Evidence yielding motivations for further research is paid special attention to. In section 6 a summary is given, together with directions for future research. The appendix contains a proof of our model being arbitrage-free, together with some technical derivations related to the state space model.

2. Call Option Valuation in Multi-Factor Affine Stochastic Volatility Option Pricing Models

2.1 The general setting of the model

The model under \mathbb{P}

Consider a financial market in which a risky stock S is traded that pays dividends at a continuous rate. Trading takes place in continuous time. The calendar time is denoted by t with $t \geq 0$ and is measured in years. There are no market imperfections in the sense that there are no transaction costs, taxes or any short sale restrictions. Furthermore the market is assumed to be competitive in the sense that a single investor cannot influence prices by his individual trades. Besides investing in the stock, investors can deposit part of their wealth in a money market account (or invest in a so-called cash bond) B that evolves deterministically over time. Its price process $\{B_t; t \geq 0\}$ is characterized by

$$dB_t = r_t B_t dt \Leftrightarrow B_t = B_0 \exp\left(\int_0^t r_s ds\right), \quad (1)$$

in which $\{r_t; t \geq 0\}$ represents the deterministic short rate process. Denote the stock price process by $\{S_t; t \geq 0\}$. The stock pays a continuously compounded dividend yield of q_t per annum at time t , such that the dividend payment in the time interval $[t, t + dt]$ of infinitesimal length equals

$$q_t S_t dt. \quad (2)$$

Under the empirical or real-world measure \mathbb{P} , the stock price follows

$$dS_t = \mu S_t dt + \sigma_t S_t dW_{S,t}, \quad (\mathbb{P}) \quad (3)$$

where μ is a constant and $\{\sigma_t; t \geq 0\}$ is the volatility process. The stock volatility varies randomly over time. In particular, the volatility is assumed to be driven by n possibly correlated, latent factors $\mathbf{x}' = (x_1, \dots, x_n)$ in an affine way,

$$\sigma_t^2 = \delta_0 + \boldsymbol{\delta}' \mathbf{x}_t, \quad (4)$$

where δ_0 and $(n \times 1) \boldsymbol{\delta} = (\delta_1, \dots, \delta_n)'$ are positively valued. Under \mathbb{P} the latent factors evolve according to stationary mean-reverting Markov diffusions,

$$d\mathbf{x}_t = \mathbf{K}(\boldsymbol{\theta} - \mathbf{x}_t)dt + \boldsymbol{\Sigma} \boldsymbol{\Lambda}_t d\mathbf{W}_{x,t}, \quad (\mathbb{P}) \quad (5)$$

where $(n \times 1) \boldsymbol{\theta}$ is the mean of the factors, $(n \times n) \mathbf{K}, \boldsymbol{\Sigma}$ are matrices of constants, and $\boldsymbol{\Lambda}_t$ is a diagonal matrix given by

$$\boldsymbol{\Lambda}_t = \text{diag}(\sqrt{\alpha_1 + \boldsymbol{\beta}_1' \mathbf{x}_t}, \dots, \sqrt{\alpha_n + \boldsymbol{\beta}_n' \mathbf{x}_t}), \quad (6)$$

in which $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)'$ and $\boldsymbol{\beta}_i = (\beta_{i1}, \dots, \beta_{in})'$, $i = 1, \dots, n$ are $(n \times 1)$ vectors of positive real valued constants. This discussion assumes the dynamics above to be well defined, in the sense that the volatility is nonnegative and $\alpha_i + \boldsymbol{\beta}_i' \mathbf{x}_t$ is nonnegative for all i and t . The matrix \mathbf{K} governs the speed of adjustment of the latent factors towards their mean $\boldsymbol{\theta}$.

In this financial market, uncertainty is resolved by the $(n+1)$ -dimensional standard Brownian motion process $\{\mathbf{W}_t; t \geq 0\}$ under \mathbb{P} , given by

$$\mathbf{W}_t = (W_{S,t}, \mathbf{W}_{x,t}')'; \quad \mathbf{W}_{x,t} = (W_{1t}, \dots, W_{nt})'. \quad (\mathbb{P}) \quad (7)$$

Its natural filtration is denoted by $\{\mathcal{F}_t; t \geq 0\}$. Notice that the Brownian motion driving the stock price is assumed to be independent of the Brownian motions driving the latent factors.

The model under \mathbb{Q}

Besides investing in the stock and the money market account, investors can trade in derivative assets like European call and put options and forward contracts written on the stock. In order to be able to obtain a fair price (i.e., a price that excludes arbitrage in this market) for the options, we also need the stochastic processes followed by the stock price and the unobservable factors under the risk-neutral probability measure \mathbb{Q} . However, as the current market setting is incomplete in the sense that not all derivative securities can be hedged for the full hundred percent by a dynamic self-financing trading strategy in the underlying stock and the bond, this measure is not unique. Moreover, each different measure will in general lead to different option prices. Hence, the condition of no-arbitrage alone is not sufficient to generate unique derivative prices. In order to obtain unique prices it is necessary to make additional assumptions with respect to the market prices of risk associated with the Brownian motions driving the factors. As these factors in turn drive the stock volatility, we might call these prices of risk collectively the *market price of volatility risk*.

In particular, to rule out arbitrage opportunities, the market price of risk belonging to the stock, $\gamma_{S,t}$, should equal

$$\gamma_{S,t} = \frac{\mu + q_t - r_t}{\sigma_t}. \quad (8)$$

Refer to the appendix for a proof. It equals the risk premium on the stock expressed per unit of risk -as measured by its standard deviation, like a Sharpe ratio. Following Duffie and Kan (1996), de Jong (2000), Dai and Singleton (2000) and many others with them, we model the market price of risk for factor x_i , denoted by γ_{it} , as being proportional to its instantaneous standard deviation,

$$\gamma_{it} = \gamma_i \sqrt{\alpha_i + \boldsymbol{\beta}_i' \mathbf{x}_t}, \quad (9)$$

in which $\gamma_i \in \mathbb{R}$. The market price of volatility risk may thus be represented by the vector $\mathbf{y}_{x,t} = (\gamma_{1t}, \dots, \gamma_{nt})'$ given by

$$\mathbf{y}_{x,t} = \boldsymbol{\Lambda}_t \boldsymbol{\gamma}, \quad (10)$$

where $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_n)'$. From Girsanov's theorem⁶, the change of measure from \mathbb{P} to \mathbb{Q} is then governed by the transformation

$$d\tilde{\mathbf{W}}_t = d\mathbf{W}_t + \mathbf{y}_t dt, \quad (11)$$

where $\mathbf{y}_t = (\gamma_{S,t}, \mathbf{y}_{x,t})'$. Here, $\{\tilde{\mathbf{W}}_t; t \geq 0\}$ with $\tilde{\mathbf{W}}_t = (\tilde{W}_{S,t}, \tilde{\mathbf{W}}_{x,t})'$ is an $(n+1)$ -dimensional standard Brownian motion under the risk-neutral measure \mathbb{Q} that has the same filtration $\{\mathcal{F}_t; t \geq 0\}$ as the \mathbb{P} -Brownian motion $\{\mathbf{W}_t; t \geq 0\}$. Given this change of measure, the stock price follows under \mathbb{Q} ,

$$dS_t = (r_t - q_t)S_t dt + \sigma_t S_t d\tilde{W}_{S,t}. \quad (\mathbb{Q}) \quad (12)$$

One advantage of the assumed form for the market price of volatility risk is the fact that it delivers the same type of mean-reverting SDE for the factors under \mathbb{Q} as under \mathbb{P} . Although the volatility function remains the same, the speed of adjustment towards the mean, and the mean itself differ under \mathbb{P} and \mathbb{Q} . Some algebraic manipulations show that under \mathbb{Q} , the factors obey the following SDE:

$$d\mathbf{x}_t = \tilde{\mathbf{K}}(\tilde{\boldsymbol{\theta}} - \mathbf{x}_t)dt + \boldsymbol{\Sigma} \boldsymbol{\Lambda}_t d\tilde{\mathbf{W}}_{x,t}, \quad (\mathbb{Q}) \quad (13)$$

where

$$\tilde{\mathbf{K}} \equiv \mathbf{K} + \boldsymbol{\Sigma} \boldsymbol{\Gamma} \boldsymbol{\beta}' \quad (14)$$

$$\tilde{\boldsymbol{\theta}} \equiv \tilde{\mathbf{K}}^{-1}(\mathbf{K}\boldsymbol{\theta} - \boldsymbol{\Sigma} \boldsymbol{\Gamma} \boldsymbol{\alpha}),$$

and

$$(n \times 1) \boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)', \quad (n \times n) \boldsymbol{\beta} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_n), \quad (n \times n) \boldsymbol{\Gamma} = \text{diag}(\gamma_1, \dots, \gamma_n),$$

and where we assume the inverse of the matrix $\tilde{\mathbf{K}}$ to exist. This completes the description of the multi-factor affine SV option pricing model.

⁶ For these prices of risk the Novikov condition for this theorem to hold is satisfied.

2.2 Call option valuation

Given the general set-up just discussed, in this section we aim at deriving the time- t value of a tradable European call option C written on the stock S , having strike price K and maturity $T > t$.

The Fundamental Theorem of Asset Pricing⁷ says that the absence of arbitrage opportunities is equivalent to the existence of at least one equivalent martingale measure \mathbb{Q} , under which the relative prices of all *tradable* securities (i.e., prices expressed in terms of the price of some numéraire asset which we take to be the cash bond), follow \mathbb{Q} -martingale processes. As our model is arbitrage-free, this theorem immediately pins down the price of the European call to

$$C_t = B_t \mathbb{E}_{\mathbb{Q}}[B_T^{-1} C_T | \mathcal{F}_t] = \exp\left(-\int_t^T r_s ds\right) \mathbb{E}_{\mathbb{Q}}[\max\{0, S_T - K\} | \mathcal{F}_t], \quad (15)$$

where $C_T = \max\{0, S_T - K\}$ is the payoff of the call at maturity. In order to compute this expectation, the conditional distribution of S_T under \mathbb{Q} is needed. It follows from applying Itô's lemma to derive the SDE for $\ln S_t$, then summing the increments and finally taking exponents to yield

$$S_T = S_t \exp\left[\left(\bar{r} - \bar{q} - \frac{1}{2}\bar{\sigma}^2\right)\tau + \int_t^T \sigma_u d\tilde{W}_{S,u}\right], \quad (\mathbb{Q}) \quad (16)$$

where we define the time to maturity τ , the average interest rate (\bar{r}), average dividend yield (\bar{q}), and average integrated variance ($\bar{\sigma}^2$), all over the remaining life of the call, by, respectively,

$$\tau \equiv T - t, \quad \bar{r} \equiv \frac{1}{\tau} \int_t^T r_u du, \quad \bar{q} \equiv \frac{1}{\tau} \int_t^T q_u du, \quad \bar{\sigma}^2 \equiv \frac{1}{\tau} \int_t^T \sigma_u^2 du. \quad (17)$$

Recall the assumption of independence between the volatility driving Brownian motions and the Brownian motion driving the stock price. Therefore, conditioning on the sample path of $\tilde{\mathbf{W}}_{x,u}$ for $t \leq u \leq T$ implies that the volatility path is known, such that the Itô integral in the previous formula is conditionally normally distributed,

$$\int_t^T \sigma_u d\tilde{W}_{S,u} | \{\tilde{\mathbf{W}}_{x,u}\} \sim \mathcal{N}\left(0, \int_t^T \sigma_u^2 du\right), \quad (\mathbb{Q}) \quad (18)$$

where $\{\tilde{\mathbf{W}}_{x,u}\}$ is shorthand notation for $\{\tilde{\mathbf{W}}_{x,u}; t \leq u \leq T\}$. But this implies that $S_T | \mathcal{F}_t, \{\tilde{\mathbf{W}}_{x,u}\}$ is lognormally distributed. Hence, we may write

$$S_T | \mathcal{F}_t, \{\tilde{\mathbf{W}}_{x,u}\} = S_t \exp\left[\left(\bar{r} - \bar{q} - \frac{1}{2}\bar{\sigma}^2\right)\tau + \bar{\sigma}\sqrt{\tau} \varepsilon\right], \quad (\mathbb{Q}) \quad (19)$$

in which ε is standard normally distributed under \mathbb{Q} . Now, express the conditional call payoff in terms of ε to get

⁷ Harrison and Kreps (1979), Harrison and Pliska (1981), Delbaen and Schachermayer (1994).

$$\max\{0, S_T - K\} | \mathcal{F}_t, \{\tilde{\mathbf{W}}_{x,u}\} = \begin{cases} S_t \exp\left[(\bar{r} - \bar{q} - \frac{1}{2}\bar{\sigma}^2)\tau + \bar{\sigma}\sqrt{\tau}\varepsilon\right] - K, & \text{if } \varepsilon > -d_2 \\ 0, & \text{if } \varepsilon \leq -d_2 \end{cases}$$

where

$$d_2 \equiv d_1 - \bar{\sigma}\sqrt{\tau}, \quad d_1 \equiv \frac{\ln \frac{S_t}{K} + (\bar{r} - \bar{q} + \frac{1}{2}\bar{\sigma}^2)\tau}{\bar{\sigma}\sqrt{\tau}}. \quad (20)$$

Given these results, we obtain

$$\begin{aligned} & \mathbb{E}_{\mathbb{Q}}[\max\{0, S_T - K\} | \mathcal{F}_t, \{\tilde{\mathbf{W}}_{x,u}\}] \\ &= S_t \exp[(\bar{r} - \bar{q})\tau] \int_{-d_2}^{\infty} \frac{1}{\sqrt{2\pi}} \exp[-\frac{1}{2}(\varepsilon - \bar{\sigma}\sqrt{\tau})^2] d\varepsilon - K \Phi(d_2) \\ &= S_t \exp[(\bar{r} - \bar{q})\tau] \Phi(d_1) - K \Phi(d_2), \end{aligned} \quad (21)$$

where $\Phi(\cdot)$ denotes the cumulative standard Gaussian distribution function, and where the last equality follows after making the change of variable towards $\xi \equiv \varepsilon - \bar{\sigma}\sqrt{\tau}$ and performing some further manipulations.

The call price at time t can subsequently be obtained by making use of the law of iterated expectations:

$$\begin{aligned} C_t &= \exp(-\bar{r}\tau) \mathbb{E}_{\mathbb{Q}}\left(\mathbb{E}_{\mathbb{Q}}[\max\{0, S_T - K\} | \mathcal{F}_t, \{\tilde{\mathbf{W}}_{x,u}\}] | \mathcal{F}_t\right) \\ &= \mathbb{E}_{\mathbb{Q}}[S_t \exp(-\bar{q}\tau) \Phi(d_1) - K \exp(-\bar{r}\tau) \Phi(d_2) | \mathcal{F}_t] \\ &= \mathbb{E}_{\mathbb{Q}}[BS(S_t, K, \tau, \bar{r}, \bar{q}, \bar{\sigma}^2) | \mathcal{F}_t], \end{aligned} \quad (22)$$

where

$$BS(S_t, K, \tau, \bar{r}, \bar{q}, \bar{\sigma}^2) \equiv S_t \exp(-\bar{q}\tau) \Phi(d_1) - K \exp(-\bar{r}\tau) \Phi(d_2) \quad (23)$$

stands for the conventional Black-Scholes call price adjusted for continuous dividend payments, evaluated in the arguments $S_t, K, \tau, \bar{r}, \bar{q}$ and $\bar{\sigma}$.

If F_t represents the fair time- t forward price of the forward contract written on S that has the same maturity τ as the call option, i.e.,

$$F_t = S_t \exp[(\bar{r} - \bar{q})\tau], \quad (24)$$

then the call valuation formula can conveniently be rewritten as

$$C_t = \mathbb{E}_{\mathbb{Q}}[BS(F_t, K, \tau, \bar{r}, \bar{\sigma}^2) | \mathcal{F}_t] \quad (25)$$

where

$$BS(F_t, K, \tau, \bar{r}, \bar{\sigma}^2) \equiv \exp(-\bar{r}\tau) [F_t \Phi(d_1) - K \Phi(d_2)] \quad (26)$$

$$d_1 = \frac{\ln \frac{F_t}{K} + \frac{1}{2}\bar{\sigma}^2\tau}{\bar{\sigma}\sqrt{\tau}}, \quad d_2 = d_1 - \bar{\sigma}\sqrt{\tau}. \quad (27)$$

The advantage of this latter expression is that the average dividend yield does not play a role in this formula, which may be hard to estimate in practice.

To conclude, in the multi-factor affine SV option pricing model, the call price is obtained as the expectation of the Black-Scholes call price, where the expectation is taken over all paths the volatility may possibly assume over the remaining maturity of the option, under the risk-neutral measure. Notice that although we allow for a much more general SV option pricing model than Hull and White do in their 1987 paper, a similar type of valuation formula results. The reason being that the volatility process is independent of the stock price process, and the fact that the stock follows a geometric Brownian motion type SDE under Q .

3. The State Space Framework

3.1 Motivating the use of the state space approach

To value a call option the Q -expectation of the Black-Scholes price evaluated in the integrated variance needs to be calculated. As no explicit analytical expression exists for this expectation, one typically relies on Monte Carlo simulation. This method assumes the parameters of the model to be known. Clearly, these are generally unknown and must first be estimated from available stock and option price data. The estimation methods commonly applied are simulation-based and therefore computationally intensive. The most prominent method is probably the Efficient Method of Moments (EMM) due to Gallant and Tauchen (1996). For a recent application in the current context see Chernov and Ghysels (2000). However EMM is not easy to implement. Besides, although EMM has desirable asymptotic properties, accumulating evidence from empirical studies indicates that the finite sample behavior of this estimation method can be very poor. Chernov and Ghysels (2000) for example mention that 'the precision of the estimates is very poor [..]' and that their result is consistent with previous findings in, for example, Gallant, Hsieh and Tauchen (1997). As another example, Duffee and Stanton (2001) find that 'EMM behaves extremely poorly in samples of the size and type usual in term structure estimation [..]'.

One of the aims of this paper is to propose a different technique that facilitates parameter estimation in SV option pricing models and that circumvents Monte Carlo simulation. The method we suggest is based on the Kalman filter and smoother and fits into the linear state space framework. Although Kalman filter based estimation methods are commonly employed in the term structure literature to estimate a wide range of interest rate models, they have not been applied to our type of model yet⁸ (at least to the best of our knowledge). The main reason probably being that it is not clear at all from the outset how the state space approach could be used in this setting. This is in rather sharp contrast to term structure models, where it is often a fairly natural way to estimate parameters. And indeed, the inspiration for our method

⁸ Nevertheless, as mentioned in the introduction, Kalman filter methods have been used extensively for estimation of SV models based on stock return data only, but not on options data. See, e.g., Harvey, Ruiz and Shephard (1994), Kim, Shephard and Chib (1998) and Sandmann and Koopman (1998).

came to mind after having read some articles on the success story of affine term structure models introduced by Duffie and Kan (1996), such as Lund (1997), de Jong (2000), Dai and Singleton (2000) and Duffie and Stanton (2001).

Using the state space methodology as a way for estimating parameters of SV option pricing models has a number of advantages. First, this framework is ultimately suited for exploiting time-series and cross-sectional information in a panel of call option price series simultaneously, besides extracting information inherent in the underlying stock price series. With a method like EMM this is probably much harder. Second, as state space models are analyzed using the Kalman filter, this filter readily delivers the filtered latent volatility series as a by-product of the estimation procedure in a natural way. In contrast, EMM requires besides parameter estimation another method called reprojection to be able to extract the volatility series. Third, the state space methodology has proven to be a rather robust estimation method in the term structure literature, see e.g. Lund (1997), de Jong (2000) and Duffie and Stanton (2001). Convergence to optimal parameter values is often rapidly obtained. Kalman filter Quasi Maximum Likelihood (QML) estimation methods tend to perform fairly well in finite samples, even though some of the different sub-methods deliver inconsistent estimates. The papers just mentioned contain Monte Carlo evidence showing that this inconsistency does not seem to be very severe in small samples.

Having motivated the possible benefits of the use of the state space methodology in the current context, we briefly review this framework in the next section. For an extensive discussion see, e.g., Hamilton (1994) or Durbin and Koopman (2001).

3.2 A brief review of the linear state space representation, the Kalman filter and smoother

The linear state space representation

We consider the following state space representation of a dynamic system, which is particularly suited for our needs. The observation- or measurement equation reads

$$\mathbf{y}_t = \mathbf{a}_t + \mathbf{H}_t' \boldsymbol{\xi}_t + \mathbf{w}_t. \quad (28)$$

Here, $(m \times 1) \mathbf{y}_t$ is a vector of variables observed at time $t = 1, \dots, T$, where time is measured in some time unit, that can be described in terms of a possibly unobserved $(r \times 1)$ vector $\boldsymbol{\xi}_t$ known as the state vector containing r state variables, and an $(m \times 1)$ vector white noise error term \mathbf{w}_t having properties

$$\mathbb{E}[\mathbf{w}_t] = \mathbf{0}; \quad \mathbb{E}[\mathbf{w}_t \mathbf{w}_s'] = \begin{cases} \mathbf{R}, & t = s \\ \mathbf{0}, & t \neq s \end{cases}. \quad (29)$$

The transition- or state equation describes the evolution of the state variables over time. They are assumed to evolve as a VAR(1) process. It reads

$$\boldsymbol{\xi}_{t+1} = \mathbf{F} \boldsymbol{\xi}_t + \mathbf{v}_{t+1}; \quad \mathbb{E}[\mathbf{v}_{t+1}] = \mathbf{0}; \quad \mathbb{E}[\mathbf{v}_{t+1} \mathbf{v}_{s+1}'] = \begin{cases} \mathbf{Q}, & t = s \\ \mathbf{0}, & t \neq s \end{cases}; \quad t = 1, \dots, T \quad (30)$$

$$\mathbb{E}[\mathbf{v}_{t+1} \boldsymbol{\xi}_1'] = \mathbf{0},$$

in which \mathbf{v}_{t+1} represents the innovation in the state at time $t+1$. The innovation series $\{\mathbf{v}_{t+1}\}$ behaves like white noise. The matrices $(mxm)\mathbf{R}$, $(rxr)\mathbf{F}$ and $(rxr)\mathbf{Q}$ are parameter matrices. The $(mx1)$ vector \mathbf{a}_t and the (mxr) matrix \mathbf{H}_t' contain parameters that are allowed to change over time, although solely in a deterministic (non-random) way. The disturbances \mathbf{w}_t and \mathbf{v}_{t+1} are assumed to be uncorrelated at all lags,

$$\text{cov}[\mathbf{w}_t, \mathbf{v}_{s+1}] = \mathbb{E}[\mathbf{w}_t \mathbf{v}_{s+1}'] = \mathbf{0} \quad \forall t, s = 1, \dots, T. \quad (31)$$

Finally, it is assumed that the initial state $\boldsymbol{\xi}_1$ is uncorrelated with the series $\{\mathbf{v}_{t+1}\}$. Given these assumptions, it follows that the state innovation \mathbf{v}_{t+1} is uncorrelated with lagged values of the state. That is $\mathbb{E}[\mathbf{v}_{t+1} \boldsymbol{\xi}_{t+1-s}'] = \mathbf{0}$; $s = 1, \dots, t$.

The Kalman filter and smoother

The representation above can be analyzed by the Kalman filter and smoother, assuming that the numerical values of \mathbf{F} , \mathbf{Q} , \mathbf{a}_t , \mathbf{H}_t and \mathbf{R} are known. The Kalman filter may be motivated as an algorithm for calculating linear least squares forecasts of the state vector on the basis of data observed through date t , that is

$$\hat{\boldsymbol{\xi}}_{t+1|t} \equiv \hat{\mathbb{E}}[\boldsymbol{\xi}_{t+1} | \mathcal{Y}_t], \quad (32)$$

where $\mathcal{Y}_t \equiv (\mathbf{y}_1', \dots, \mathbf{y}_t')$ and $\hat{\mathbb{E}}[\boldsymbol{\xi}_{t+1} | \mathcal{Y}_t]$ denotes the linear projection of $\boldsymbol{\xi}_{t+1}$ on \mathcal{Y}_t and a constant. The Kalman filter calculates these forecasts recursively, generating $\hat{\boldsymbol{\xi}}_{1|0}$, $\hat{\boldsymbol{\xi}}_{2|1}$, ..., $\hat{\boldsymbol{\xi}}_{T+1|T}$ in succession. Associated with each of these forecasts is a mean squared error (MSE) or variance matrix $\mathbf{P}_{t+1|t}$ given by

$$\mathbf{P}_{t+1|t} \equiv \mathbb{E}[(\boldsymbol{\xi}_{t+1} - \hat{\boldsymbol{\xi}}_{t+1|t})(\boldsymbol{\xi}_{t+1} - \hat{\boldsymbol{\xi}}_{t+1|t})']. \quad (33)$$

Provided that the eigenvalues of \mathbf{F} are inside the unit circle such that the process for $\boldsymbol{\xi}_t$ is covariance stationary, the Kalman filter can be started with the unconditional mean and variance of $\boldsymbol{\xi}_1$,

$$\hat{\boldsymbol{\xi}}_{1|0} = \mathbb{E}[\boldsymbol{\xi}_1] = \mathbf{0} \quad (34)$$

$$\text{vec}(\mathbf{P}_{1|0}) = \text{vec}(\mathbb{E}[(\boldsymbol{\xi}_1 - \hat{\boldsymbol{\xi}}_{1|0})(\boldsymbol{\xi}_1 - \hat{\boldsymbol{\xi}}_{1|0})']) = \text{vec}(\text{var}[\boldsymbol{\xi}_1]) = [\mathbf{I}_{r^2} - (\mathbf{F} \otimes \mathbf{F})]^{-1} \text{vec}(\mathbf{Q}).$$

We then iterate on

$$\hat{\boldsymbol{\xi}}_{t+1|t} = \mathbf{F} \hat{\boldsymbol{\xi}}_{t|t-1} + \mathbf{F} \mathbf{P}_{t|t-1} \mathbf{H}_t' (\mathbf{H}_t \mathbf{P}_{t|t-1} \mathbf{H}_t' + \mathbf{R})^{-1} (\mathbf{y}_t - \mathbf{a}_t - \mathbf{H}_t' \hat{\boldsymbol{\xi}}_{t|t-1}) \quad (35)$$

$$\mathbf{P}_{t+1|t} = \mathbf{F} [\mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} \mathbf{H}_t' (\mathbf{H}_t \mathbf{P}_{t|t-1} \mathbf{H}_t' + \mathbf{R})^{-1} \mathbf{H}_t \mathbf{P}_{t|t-1}] \mathbf{F}' + \mathbf{Q}; \quad t = 1, \dots, T.$$

The Kalman filter is mainly used for producing a forecast of the future state, given current data. In contrast, the Kalman smoother is particularly suited for obtaining an estimate of the state, given *all* observations, i.e., for $\hat{\boldsymbol{\xi}}_{t+1|T} \equiv \hat{\mathbb{E}}[\boldsymbol{\xi}_{t+1} | \mathcal{Y}_T]$, with associated MSE matrix $\mathbf{P}_{t+1|T} \equiv \mathbb{E}[(\boldsymbol{\xi}_{t+1} - \hat{\boldsymbol{\xi}}_{t+1|T})(\boldsymbol{\xi}_{t+1} - \hat{\boldsymbol{\xi}}_{t+1|T})']$. These smoothed estimates of the state can be obtained from a backwards recursion that uses the output from the Kalman filter. See, for example, Hamilton (1994) for details on this.

Estimation

In practice the numerical values of \mathbf{F} , \mathbf{Q} , \mathbf{a}_t , \mathbf{H}_t and \mathbf{R} are typically unknown and must be estimated from the data \mathcal{Y}_T at hand. If the initial state $\boldsymbol{\xi}_1$ and the series $\{\mathbf{w}_t, \mathbf{v}_{t+1}\}$ are multivariate Gaussian, then the distribution of \mathbf{y}_t conditional on \mathcal{Y}_{t-1} is Gaussian:

$$\mathbf{y}_t | \mathcal{Y}_{t-1} \sim \mathcal{N}[\mathbf{a}_t + \mathbf{H}_t' \hat{\boldsymbol{\xi}}_{t|t-1}; \mathbf{H}_t' \mathbf{P}_{t|t-1} \mathbf{H}_t + \mathbf{R}], \quad (36)$$

from which the sample loglikelihood is easily constructed. The loglikelihood can subsequently be maximized numerically with respect to the unknown parameters in the matrices \mathbf{F} , \mathbf{Q} , \mathbf{a}_t , \mathbf{H}_t and \mathbf{R} . In the absence of sufficient restrictions on these matrices the parameters are unidentified. If the disturbances $\{\mathbf{w}_t\}$ and $\{\mathbf{v}_{t+1}\}$ are non-Gaussian, the Kalman filter can still be used to calculate the linear projection of \mathbf{y}_{t+s} on past observable variables. Moreover, we can still write down the same loglikelihood as before and maximize it with respect to the unknown parameters, even for non-Gaussian systems. This quasi maximum likelihood (QML) estimation procedure will still yield consistent and asymptotically normal estimates of the parameters, provided that some mild regularity conditions are satisfied (see Watson (1989)). So far this review.

4. Deriving the State Space Representation of our Model

4.1 The equations from the option pricing model collected

In this part the linear state space representation of the affine SV option pricing model is derived. For convenience we first repeat the basic equations of the option pricing model. From now on we concentrate on the model with diagonal \mathbf{K} , which we denote by $\mathbf{K}_d = \text{diag}(k_1, \dots, k_n)$. These equations are:

$$\text{Stock price:} \quad dS_t = \mu S_t dt + \sigma_t S_t dW_{S,t} \quad (\text{P}) \quad (37)$$

$$\text{Variance:} \quad \sigma_t^2 = \delta_0 + \boldsymbol{\delta}' \mathbf{x}_t \quad (38)$$

$$\text{Latent state:} \quad d\mathbf{x}_t = \mathbf{K}_d(\boldsymbol{\theta} - \mathbf{x}_t)dt + \boldsymbol{\Sigma} \boldsymbol{\Lambda}_t d\mathbf{W}_{x,t} \quad (\text{P}) \quad (39)$$

$$d\mathbf{x}_t = \tilde{\mathbf{K}}(\tilde{\boldsymbol{\theta}} - \mathbf{x}_t)dt + \boldsymbol{\Sigma} \boldsymbol{\Lambda}_t d\tilde{\mathbf{W}}_{x,t} \quad (\text{Q}) \quad (40)$$

$$\tilde{\mathbf{K}} \equiv \mathbf{K}_d + \boldsymbol{\Sigma} \boldsymbol{\Gamma} \boldsymbol{\beta}' , \tilde{\boldsymbol{\theta}} \equiv \tilde{\mathbf{K}}^{-1}(\mathbf{K}_d \boldsymbol{\theta} - \boldsymbol{\Sigma} \boldsymbol{\Gamma} \boldsymbol{\alpha}) \quad (41)$$

$$\boldsymbol{\Lambda}_t = \text{diag}(\sqrt{\alpha_1 + \boldsymbol{\beta}_1' \mathbf{x}_t}, \dots, \sqrt{\alpha_n + \boldsymbol{\beta}_n' \mathbf{x}_t}) \quad (42)$$

$$\text{Call premium:} \quad C_t = \mathbb{E}_{\mathbb{Q}}[BS(S_t, K, \tau, \bar{r}, \bar{q}, \bar{\sigma}^2) | \mathcal{F}_t] \quad (43)$$

Given this continuous-time model, our object is to translate the model into the discrete-time state space representation discussed above, which can be used for estimation purposes. We assume the empirical data to consist of T daily observations of stock returns and a panel of call option prices. Time t will be measured in trading years. As a year is assumed to consist of 260 trading days, the timing of the data points will be denoted by $t = \Delta t, 2\Delta t, 3\Delta t, \dots, T\Delta t$ with $\Delta t = 1/260$. Each model

equation will be considered separately in the sub-sections below. After that a summary is given.

4.2 Tackling the stock price SDE

Let us first focus on the SDE for the stock price. Unfortunately, an exact discretization does not exist. From its Euler discretization⁹ over the interval $[t, t + \Delta t]$ we obtain

$$r_{t+\Delta t} = \mu\Delta t + \sigma_t \eta_{t+\Delta t}, \quad (44)$$

where the stock return $r_{t+\Delta t}$ and the Brownian increment $\eta_{t+\Delta t}$ are defined as

$$r_{t+\Delta t} \equiv \frac{S_{t+\Delta t} - S_t}{S_t}; \quad \eta_{t+\Delta t} \equiv W_{S,t+\Delta t} - W_{S,t} \sim \text{i.i.d. } \mathcal{N}(0, \Delta t). \quad (45)$$

Given the filtration \mathcal{F}_t , the returns are conditionally normally distributed:

$$r_{t+\Delta t} | \mathcal{F}_t \sim \mathcal{N}(\mu\Delta t, \sigma_t^2 \Delta t). \quad (46)$$

Notice that the series $\{\sigma_t^2\}$ represents the *conditional* variance series (per annum) of the stock returns, which varies over time in a random way. The *unconditional* per annum variance of the stock returns is given by $\mathbb{E}[\sigma_t^2]$. In the present analysis, we are ultimately interested in the volatility. As volatility clustering is associated with temporal dependence in second order central moments, our focus is on squared returns in deviation from their mean instead of the raw returns themselves. Consider the equation

$$(r_{t+\Delta t} - \mu\Delta t)^2 = \mathbb{E}[(r_{t+\Delta t} - \mu\Delta t)^2 | \mathcal{F}_t] + \bar{\omega}_{t+\Delta t}, \quad (47)$$

where, by construction, the error term $\bar{\omega}_{t+\Delta t}$ has mean zero. Some rewriting yields

$$\frac{1}{\Delta t} (r_{t+\Delta t} - \mu\Delta t)^2 = \delta_0 + \boldsymbol{\delta}' \mathbf{x}_t + \omega_{t+\Delta t}, \quad (48)$$

where $\omega_{t+\Delta t} \equiv \bar{\omega}_{t+\Delta t} / \Delta t$. Notice that this equation is linear in the latent factors. What remains to be checked is the statistical properties of the series $\{\omega_{t+\Delta t}\}$. To be able to use the estimation method based on the Kalman filter described previously, this series ought to be white noise. And it indeed is,

$$\mathbb{E}[\omega_{t+\Delta t}] = 0, \quad \mathbb{E}[\omega_{t+\Delta t} \omega_{s+\Delta t}] = \begin{cases} \sigma_\omega^2, & t = s \\ 0, & t \neq s \end{cases}. \quad (49)$$

A proof together with an explicit expression for σ_ω^2 is given in the appendix. The equation for the squared returns will serve as part of the measurement equation¹⁰.

⁹ This Euler discretization is not exact. Nevertheless we write an equality sign = instead of an 'approximately equal' sign \approx .

¹⁰ Prior to estimation, we substitute the sample average of the returns for μ to make the left hand side "observable".

4.3 Tackling the call premium formula

In practice, we also observe the call price series. The big issue is now how to extract the inherent information from this source. Looking at the call valuation formula above, it is not clear at all how to translate this expression into an equation that can serve as the second part of the measurement equation of the linear state space representation. An exact way does not exist. Below we propose an approximate method based on linearizing the Black-Scholes function in order to put the pricing formula into the desired form.

As a prelude, recall the volatility process under \mathbb{Q} ; see (38) and (40). Given this specification and invoking the results of Duffie and Kan (1996)¹¹, we know that the conditional \mathbb{Q} -expectation of the exponent of the integrated variance is an exponential-affine function of the latent factors,

$$\mathbb{E}_{\mathbb{Q}}[\exp(\int_t^T \sigma_u^2 du) | \mathcal{F}_t] = \exp[A(\tau) + \mathbf{B}(\tau)' \mathbf{x}_t], \quad (50)$$

where $A(\tau)$ is some deterministic function of the time to maturity $\tau \equiv T - t$, and $\mathbf{B}(\tau)$ is an $(n \times 1)$ deterministic vector function of τ . These functions satisfy the following system of ordinary differential equations (ODEs),

$$\begin{aligned} \frac{dA(\tau)}{d\tau} &= \tilde{\boldsymbol{\theta}}' \tilde{\mathbf{K}}' \mathbf{B}(\tau) + \frac{1}{2} \sum_{i=1}^n [\boldsymbol{\Sigma}' \mathbf{B}(\tau)]_i^2 \alpha_i + \delta_0 \\ \frac{d\mathbf{B}(\tau)}{d\tau} &= -\tilde{\mathbf{K}}' \mathbf{B}(\tau) + \frac{1}{2} \sum_{i=1}^n [\boldsymbol{\Sigma}' \mathbf{B}(\tau)]_i^2 \boldsymbol{\beta}_i + \boldsymbol{\delta}, \end{aligned} \quad (51)$$

with initial conditions $A(0) = 0$ and $\mathbf{B}(0) = \mathbf{0}$. Notice that $A(\tau)$ and $\mathbf{B}(\tau)$ are functions of the risk-neutral parameters $\tilde{\mathbf{K}}$ and $\tilde{\boldsymbol{\theta}}$. For Gaussian models (i.e., models in which the volatility function of the SDE for the factors is deterministic), we can solve for $A(\tau)$ and $\mathbf{B}(\tau)$ explicitly. For other models the system can relatively easily be solved numerically with the Runge-Kutta method for example.

4.3.1 Linearizing the call price formula

Our aim is now to incorporate these convenient “analytical” results in a smart way when deriving a possible state space equation from the call price formula. Reconsider this formula (43). Notice that the arguments of the Black-Scholes pricing function are known at time t given the filtration \mathcal{F}_t , except for the average integrated variance $\bar{\sigma}^2$ over the remaining life of the option, which is obviously random. The

¹¹ They model the short rate by $r_t = \delta_0 + \boldsymbol{\delta}' \mathbf{x}_t$ in which the latent factors \mathbf{x}_t follow the same SDE as above. Given this set up, they show that the price of a zero coupon bond is given by

$$P(t, T) = \mathbb{E}_{\mathbb{Q}}[\exp(-\int_t^T r_s ds) | \mathcal{F}_t] = \exp[A(\tau) + \mathbf{B}(\tau)' \mathbf{x}_t],$$

where $A(\tau)$ and $\mathbf{B}(\tau)$ satisfy a similar system of ODEs as above.

exponent of the integrated variance is the variable we are going to concentrate on now. In order to do so, define

$$X \equiv \int_t^T \sigma_s^2 ds, \quad Y \equiv \exp(X) = \exp\left(\int_t^T \sigma_s^2 ds\right) \quad (52)$$

The Black-Scholes price is a function $f(\cdot)$ of the integrated variance X . Given that, conditional on \mathcal{F}_t , all other arguments are known constants, we may write

$$f(X) \equiv BS(S_t, K, \tau, \bar{r}, \bar{q}, \tau^{-1}X). \quad (53)$$

It is also true that

$$f(X) = f(\ln[\exp X]) = g(\exp X) = g(Y), \quad (54)$$

where $g(\cdot) \equiv f(\ln[\cdot])$. Now, for the crux of the argument, consider a first order Taylor series expansion of the function $g(\cdot)$ around the point $Y = b$,

$$g(Y) = g(b) + (Y - b)g'(b) + HOT, \quad (55)$$

where *HOT* stands for higher order terms. Then, the theoretical call value can be written as

$$\begin{aligned} C_t &= \mathbb{E}_{\mathbb{Q}}[BS(S_t, K, \tau, \bar{r}, \bar{q}, \bar{\sigma}^2) | \mathcal{F}_t] \\ &= \mathbb{E}_{\mathbb{Q}}[g(Y) | \mathcal{F}_t] \\ &= \mathbb{E}_{\mathbb{Q}}[g(b) + (Y - b)g'(b) + HOT | \mathcal{F}_t] \\ &= g(b) - bg'(b) + \mathbb{E}_{\mathbb{Q}}[Y | \mathcal{F}_t]g'(b) + \mathbb{E}_{\mathbb{Q}}[HOT | \mathcal{F}_t] \end{aligned} \quad (56)$$

Note that $g(b)$ and $g'(b)$ are easily calculated. Rewriting yields

$$\frac{C_t - g(b) + bg'(b) - \mathbb{E}_{\mathbb{Q}}[HOT | \mathcal{F}_t]}{g'(b)} = \mathbb{E}_{\mathbb{Q}}[Y | \mathcal{F}_t] = \exp[A(\tau) + \mathbf{B}(\tau)' \mathbf{x}_t], \quad (57)$$

where the last equality follows from the assumption of an affine SV model for the stock price, and where $A(\tau)$ and $\mathbf{B}(\tau)$ satisfy the previously given system of ODEs (51). Taking logarithms yields

$$\ln\left(\frac{C_t - g(b) + bg'(b) - \mathbb{E}_{\mathbb{Q}}[HOT | \mathcal{F}_t]}{g'(b)}\right) = A(\tau) + \mathbf{B}(\tau)' \mathbf{x}_t. \quad (58)$$

This equation is linear in the unobservable factors \mathbf{x}_t .

4.3.2 Making use of the theoretical relationship in practice

Equation (58) is an exact relationship that describes how the theoretical call value is related to the latent factors. How can this equation be implemented for practical purposes, i.e., for the purpose of parameter estimation?

We linearized the Black-Scholes pricing function around a certain point b without further characterizing it. Some obvious issues naturally arise. Around which point should we linearize? And what about the quality of the linearization? Is the function $g(\cdot)$ sufficiently well behaved to be adequately approximated by a linear function? Regarding the point of linearization, b , we argue as follows. Consider the conventional Black-Scholes world with time-varying but deterministic volatility. In that case the call price equals the conventional Black-Scholes price but with the variance parameter σ^2 replaced by the average integrated variance X/τ over the remaining maturity of the option. Hence, the Black-Scholes implied variance essentially represents the average integrated variance in this setting. That is, $\sigma_{implied,t}^2 = X/\tau$. Now remember from above that we view the Black-Scholes pricing function as a function of Y , being the exponent of the integrated variance, $Y = \exp(X)$. Therefore, an arguably reasonable point of linearization could be $b_t^* \equiv \exp(\tau \sigma_{implied,t}^2)$, which is easily calculated. Notice that this linearization is performed for each t , such that the value of the point of linearization differs for each t ; it equals b_t^* . Notice that for this specific choice of b we find that

$$g(b_t^*) = g(\exp[\tau \sigma_{implied,t}^2]) = f(\tau \sigma_{implied,t}^2) = BS(S_t, K, \tau, \bar{r}, \bar{q}, \sigma_{implied,t}^2) \equiv C_t^{mkt}, \quad (59)$$

where C_t^{mkt} represents the observed market value of the call. Substitution into the theoretical relationship subsequently yields

$$\ln\left(\frac{C_t - C_t^{mkt} - \mathbb{E}_{\mathbb{Q}}[HOT | \mathcal{F}_t]}{g'(b_t^*)} + b_t^*\right) = A(\tau) + \mathbf{B}(\tau)' \mathbf{x}_t. \quad (60)$$

Some comments on the quality of this linearization will be postponed till the end of this section.

For a practical implementation of the above relationship, we next propose to replace the theoretical call value C_t with its observed counterpart C_t^{mkt} , and to neglect the higher order terms. To more or less 'compensate' for these simplifications, and the fact that a model is never a complete description of reality, we introduce noise in the form of an additive random error term v_t for which we assume that it is uncorrelated with any other variable or series. We then obtain

$$\tau \sigma_{implied,t}^2 = A(\tau) + \mathbf{B}(\tau)' \mathbf{x}_t + v_t, \quad (61)$$

since $\ln b_t^* = \tau \sigma_{implied,t}^2$ ¹². An important issue concerns the assumptions about the statistical properties of v_t , and in particular its first two moments. These cannot be derived in a straightforward way. Is it reasonable to assume that v_t has mean zero? Probably not. Instead, we assume that the mean is proportional to the time to maturity,

$$\mathbb{E}[v_t] = \mu_v \tau, \quad (62)$$

¹² Compare this equation to the often encountered equation in the term structure literature, where the product of the time to maturity and the bond yield is on the left hand side.

where μ_v is a constant. And what about its variance? Reconsider equation (61). In a deterministic volatility setting the left hand side essentially equals the integrated variance X over the interval $[t, T]$. Hence, we expect that the larger τ , the larger the value of the left hand side will be, and the larger its variance. In contrast, the variance of the right hand side variables \mathbf{x}_t is constant. Hence, it may not be too unreasonable to assume that the variance of v_t depends on τ , such that $\{v_t\}$ is heteroskedastic. In particular (and mostly for convenience), we will assume that

$$\text{var}[v_t] = \sigma_\varepsilon^2 \tau^2, \quad (63)$$

such that the standard deviation of v_t is proportional to the maturity of the call. Furthermore, we assume that the series $\{v_t\}$ is non-autocorrelated:

$$\text{cov}[v_t, v_s] = 0 \quad \forall t \neq s. \quad (64)$$

Given these assumptions equation (61) can be rewritten. Towards this end, introduce the error term

$$\varepsilon_t \equiv \frac{v_t - \mu_v \tau}{\tau}, \quad (65)$$

which is a white noise series by construction for which

$$\begin{aligned} \mathbb{E}[\varepsilon_t] &= 0; \quad \text{var}[\varepsilon_t] = \sigma_\varepsilon^2; \quad \text{cov}[\varepsilon_t, \varepsilon_s] = 0 \quad \forall t \neq s; \\ \varepsilon_t &\text{ uncorrelated with any other variable or series} \end{aligned} \quad (66)$$

These properties follow from the assumed properties of v_t . Substitution, followed by rewriting eventually yields

$$\sigma_{implied,t}^2 = \mu_v + \frac{A(\tau)}{\tau} + \frac{\mathbf{B}(\tau)'}{\tau} \mathbf{x}_t + \varepsilon_t \quad (67)$$

This equation will serve as the second part of the measurement equation of the state space model.

4.3.3 Some preliminary insight into the quality of the linearization

We now provide an example that yields some preliminary insight into the quality of the linearization of the Black-Scholes pricing function around the point $b_t^* = \exp(\tau \sigma_{implied,t}^2)$, regarded as function of Y . For simplicity, we assume the dividend yield to be zero. This function equals

$$g(Y) = S_t \Phi[d_1(Y)] - K \exp(-\bar{r}\tau) \Phi[d_2(Y)], \quad (68)$$

where

$$d_1(Y) = \frac{\ln \frac{S_t}{K} + \bar{r}\tau + \frac{1}{2} \ln Y}{\sqrt{\ln Y}}; \quad d_2(Y) = d_1(Y) - \sqrt{\ln Y}.$$

The first thing one should ask is what a reasonable interval for Y is to lie in, in practice. Obviously, Y is always larger than 1 since the integrated variance is always

larger than zero. A reasonable interval for practical values of implied volatilities may be $\sigma_{BS,implied} \in [1\%, 80\%]$ suggesting that a reasonable interval for Y could be $[\exp(0.0001\tau), \exp(0.64\tau)]$.

Now, consider for instance an in-the-money call option on a stock that is currently worth 42, having strike $K = 40$ and maturity $\tau = 0.5$ years. Suppose further that the current interest rate on T-bills maturing half a year from now equals 10%. In this case, the interval for reasonable values of Y is $[\exp(0.00005), \exp(0.32)]$ which equals $[1.00005, 1.38]$. Below the graph of $g(\cdot)$ is drawn for this interval Y .

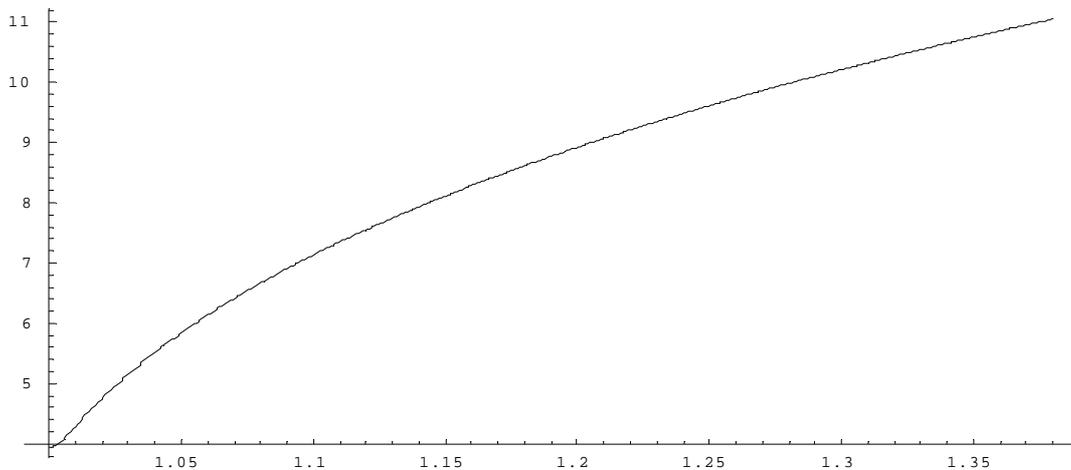


Figure 1: The function $g(Y)$ for $Y \in [1.00005, 1.38]$.

Zooming in on Y reveals that in the very close neighborhood of 1, the function behaves as shown in figure 2.

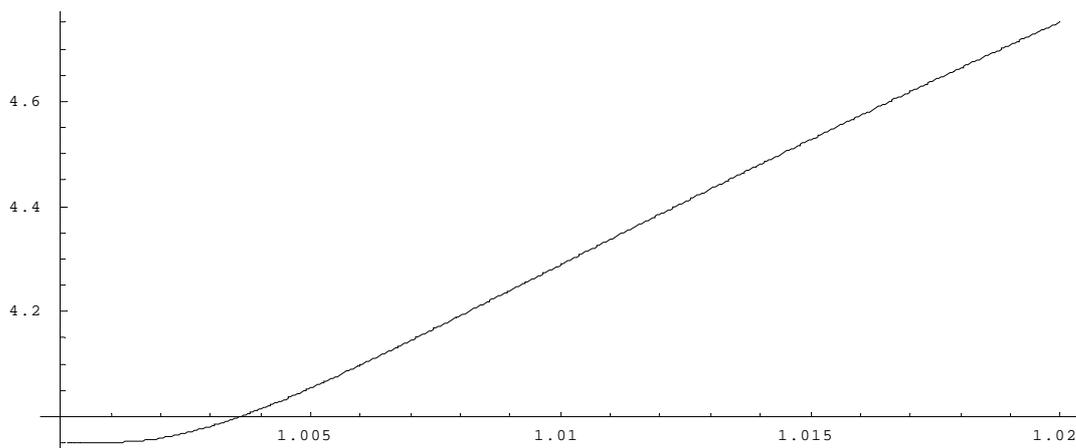


Figure 2: Zooming in on the function $g(Y)$ in the close neighborhood of 0.

Suppose that the call has a value of 5.71 such that the Black-Scholes implied volatility equals $\sigma_{BS,implied} = 30\%$. We thus linearize around $b^* = \exp(0.5 \cdot 0.3^2) = 1.046$. Now, notice that the theoretical call pricing formula (43) in the affine SV option pricing model essentially tells us to compute the call value as a “probability weighted average” of Black-Scholes values for all different values that Y can assume, $C_t = \mathbb{E}_{\mathbb{Q}}[g(Y) | \mathcal{F}_t]$. Intuitively, we expect the most “probability weight” to be put on values of Y around the point of linearization, where the error is smallest. The further away we are from the linearization point, we expect the greater errors, but the less weight these points have in the overall call value. Figure 3 draws $g(Y)$ together with its linear approximation in $b^* = 1.046$ for the interval $[1.005, 1.13]$ corresponding to $\sigma_{BS,implied} \in [10\%, 50\%]$.

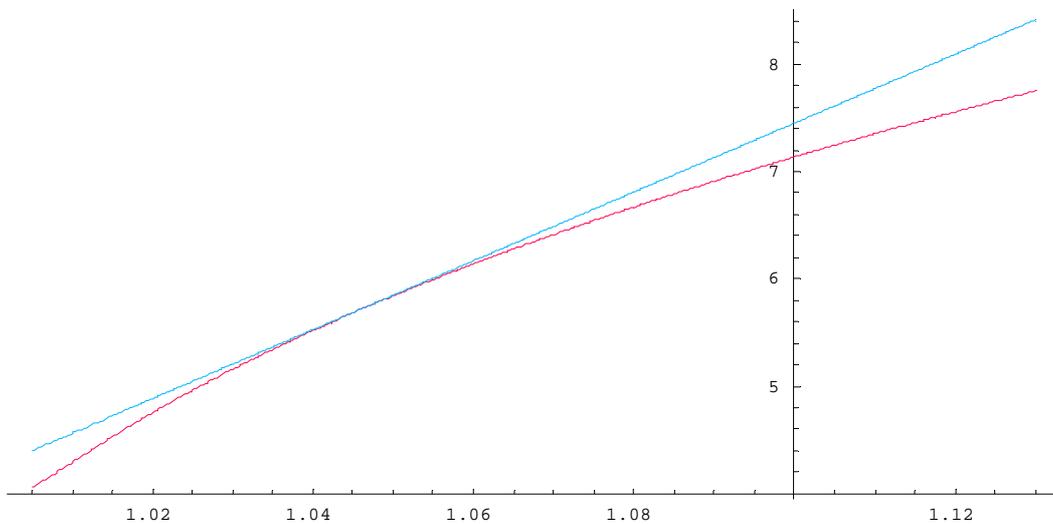


Figure 3: The function $g(Y)$ and its linear approximation in $b^* = 1.046$.

It is also instructive to consider the approximation error for these values of Y , shown underneath.

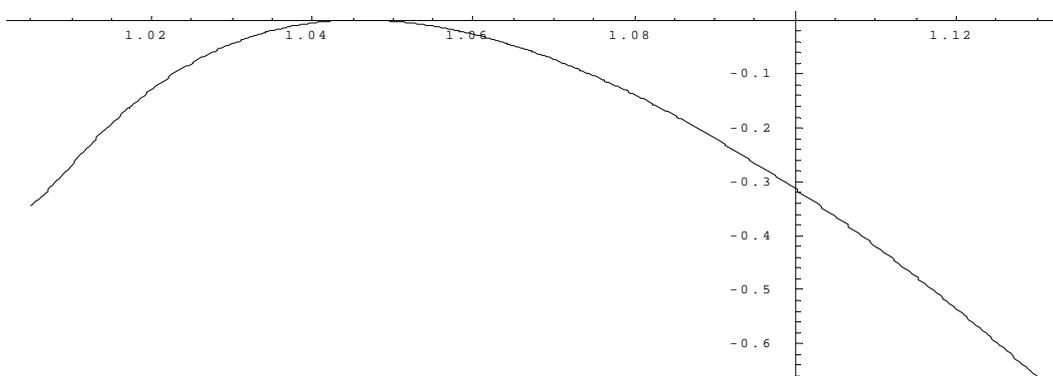


Figure 4: The approximation error between $g(Y)$ and its linear approximation in $b^* = 1.046$.

Although further investigation should make things clearer, this preliminary analysis suggests that a linear approximation may not be too bad.

4.4 Deriving the transition equation

The object in this section is to derive the transition equation of the state space model which describes the development of the state over time. In our set up the state is formed by some function of the unobservable factors. Recall their SDE under \mathbb{P} ,

$$d\mathbf{x}_t = \mathbf{K}_d(\boldsymbol{\theta} - \mathbf{x}_t)dt + \boldsymbol{\Sigma}\boldsymbol{\Lambda}_t d\mathbf{W}_{x,t}. \quad (\mathbb{P}) \quad (69)$$

To obtain a discrete-time equation from this SDE, one could consider its Euler discretization. However, fortunately an exact discretization exists. First, we express $\mathbf{x}_{t+\Delta t}$ in terms of \mathbf{x}_t . In order to do so, consider the transformation

$$\mathbf{y}_t \equiv \exp[\mathbf{K}_d t] \mathbf{x}_t, \quad (70)$$

where we define the exponent of a diagonal matrix as follows. If $(n \times n) \mathbf{A}_d = \text{diag}(a_1, \dots, a_n)$, then $\exp[\mathbf{A}_d] \equiv \text{diag}(\exp[a_1], \dots, \exp[a_n])$. By Itô's lemma,

$$d\mathbf{y}_t = \mathbf{K}_d \exp[\mathbf{K}_d t] \mathbf{x}_t dt + \exp[\mathbf{K}_d t] d\mathbf{x}_t = \exp(\mathbf{K}_d t) [\mathbf{K}_d \boldsymbol{\theta} dt + \boldsymbol{\Sigma}\boldsymbol{\Lambda}_t d\mathbf{W}_{x,t}]. \quad (71)$$

Summing the increments over the interval $[t, t + \Delta t]$ yields

$$\mathbf{y}_{t+\Delta t} = \mathbf{y}_t + \int_t^{t+\Delta t} \exp[\mathbf{K}_d u] \mathbf{K}_d \boldsymbol{\theta} du + \int_t^{t+\Delta t} \exp[\mathbf{K}_d u] \boldsymbol{\Sigma}\boldsymbol{\Lambda}_u d\mathbf{W}_{x,u}. \quad (72)$$

Now, transforming back to $\mathbf{x}_{t+\Delta t}$ by premultiplying with $\exp[-\mathbf{K}_d(t + \Delta t)]$ yields

$$\begin{aligned} \mathbf{x}_{t+\Delta t} &= \exp[-\mathbf{K}_d \Delta t] \mathbf{x}_t + \int_t^{t+\Delta t} \exp[-\mathbf{K}_d(t + \Delta t - u)] \mathbf{K}_d \boldsymbol{\theta} du + \int_t^{t+\Delta t} \exp[-\mathbf{K}_d(t + \Delta t - u)] \boldsymbol{\Sigma}\boldsymbol{\Lambda}_u d\mathbf{W}_{x,u} \\ &= \exp[-\mathbf{K}_d \Delta t] \mathbf{x}_t + (\mathbf{I}_n - \exp[-\mathbf{K}_d \Delta t]) \boldsymbol{\theta} + \int_t^{t+\Delta t} \exp[-\mathbf{K}_d(t + \Delta t - u)] \boldsymbol{\Sigma}\boldsymbol{\Lambda}_u d\mathbf{W}_{x,u}. \end{aligned} \quad (73)$$

Defining

$$\mathbf{x}_t^* \equiv \mathbf{x}_t - \boldsymbol{\theta} \quad \forall t; \quad \mathbf{u}_{t+\Delta t} \equiv \int_t^{t+\Delta t} \exp[-\mathbf{K}_d(t + \Delta t - u)] \boldsymbol{\Sigma}\boldsymbol{\Lambda}_u d\mathbf{W}_{x,u} \quad (74)$$

this equation can be simplified towards

$$\mathbf{x}_{t+\Delta t}^* = \exp[-\mathbf{K}_d \Delta t] \mathbf{x}_t^* + \mathbf{u}_{t+\Delta t}; \quad t = \Delta t, \dots, T \Delta t. \quad (75)$$

This equation already has the basic form we are looking for, if we interpret the random vector $\mathbf{u}_{t+\Delta t}$ as the error term corresponding to the state equation. Notice that the state vector $\boldsymbol{\xi}_t$ from the state space model will be formed by the latent factors in deviation from their mean, i.e. by $\mathbf{x}_t^* \equiv \mathbf{x}_t - \boldsymbol{\theta}$. What remains to be

checked however, is the statistical properties of the series $\{\mathbf{u}_{t+\Delta t}\}$, $t = \Delta t, \dots, T\Delta t$. Recalling the state space formulation (30), this series ought to be white noise. In the appendix we prove that this is indeed the case. In particular, we find

$$\mathbb{E}[\mathbf{u}_{t+\Delta t}] = \mathbf{0}; \quad \mathbb{E}[\mathbf{u}_{t+\Delta t} \mathbf{u}_{s+\Delta t}'] = \begin{cases} \mathbf{G} \odot \Sigma \mathbf{M}_d \Sigma'; & t = s \\ \mathbf{0}; & t \neq s \end{cases} \quad (76)$$

in which

$$\mathbf{M}_d \equiv \text{diag}[\alpha_1 + \boldsymbol{\beta}_1' \boldsymbol{\theta}, \dots, \alpha_n + \boldsymbol{\beta}_n' \boldsymbol{\theta}] \quad (77)$$

$$(n \times n) \mathbf{G} \text{ with } [\mathbf{G}]_{ij} = \frac{1 - \exp[-(k_i + k_j)\Delta t]}{k_i + k_j},$$

and \odot denotes the Hadamard product; i.e., element-by-element multiplication. Furthermore it is shown that the error term $\mathbf{u}_{t+\Delta t}$ is uncorrelated with lagged values of the state. Finally, it is proven that the error terms of the measurement equation and the transition equation are uncorrelated as they ought to be.

4.5 The state space model

In the previous sections we derived the individual ingredients of the discrete-time state space model extracted from our continuous-time SV option pricing model. Here, we state the resulting state space model that can be used for the purpose of parameter estimation. As having shown, the model exactly fits into the general framework discussed in section 3.2. Using the notation from there the following state space model results, where $t = \Delta t, 2\Delta t, \dots, T\Delta t$ ¹³:

$$\begin{aligned} \mathbf{y}_t &= \mathbf{a}_t + \mathbf{H}_t' \boldsymbol{\xi}_t + \mathbf{w}_t, & \mathbf{w}_t &\sim (\mathbf{0}, \mathbf{R}), \\ \boldsymbol{\xi}_{t+\Delta t} &= \mathbf{F} \boldsymbol{\xi}_t + \mathbf{v}_{t+\Delta t}, & \mathbf{v}_{t+\Delta t} &\sim (\mathbf{0}, \mathbf{Q}), \end{aligned} \quad (78)$$

with

$$\mathbf{y}_t \equiv \begin{bmatrix} \frac{1}{\Delta t} (r_{t+\Delta t} - \hat{\mu} \Delta t)^2 \\ \sigma_{implied,t}^2 \end{bmatrix}, \quad \mathbf{a}_t \equiv \begin{bmatrix} \delta_0 + \boldsymbol{\delta}' \boldsymbol{\theta} \\ \mu_v + \frac{1}{\tau} [A(\tau) + \mathbf{B}(\tau)' \boldsymbol{\theta}] \end{bmatrix},$$

$$\mathbf{H}_t' \equiv \begin{bmatrix} \boldsymbol{\delta}' \\ \frac{1}{\tau} \mathbf{B}(\tau)' \end{bmatrix}, \quad \boldsymbol{\xi}_t \equiv \mathbf{x}_t^*, \quad \mathbf{w}_t \equiv \begin{pmatrix} \omega_{t+\Delta t} \\ \varepsilon_t \end{pmatrix}, \quad \mathbf{R} = \begin{pmatrix} \sigma_\omega^2 & 0 \\ 0 & \sigma_\varepsilon^2 \end{pmatrix},$$

$$\mathbf{F} \equiv \exp[-\mathbf{K}_d \Delta t], \quad \mathbf{v}_{t+\Delta t} \equiv \mathbf{u}_{t+\Delta t}, \quad \mathbf{Q} = \mathbf{G} \odot \Sigma \mathbf{M}_d \Sigma',$$

¹³ Notice that since the state is formed by $\mathbf{x}_t^* = \mathbf{x}_t - \boldsymbol{\theta}$, the equations that constitute together the measurement equation have been rewritten in terms of this state. Notice furthermore that we repeat the main equations only.

with \mathbf{G} , \mathbf{M}_d as above and σ_ω^2 as in the appendix. Notice that although the first two moments of the error terms \mathbf{w}_t and $\mathbf{v}_{t+\Delta t}$ are known, their distribution is unknown. As explained in section 3.2, the parameters will be estimated by QML which still yields consistent and asymptotically normal estimates.

Notice that this formulation supposes that only one call option price series is analyzed besides the stock price series. Generalizing the formulation to be able to tackle more than one call simultaneously is more or less obvious. One could then assume the error terms belonging to each call equation to have its own variance and perhaps to be cross-sectionally correlated, but to be uncorrelated over time.

5. Empirical Results

5.1 Data

We examine the FTSE100 index which is based on a portfolio of 100 major U.K. stocks listed on the London Stock Exchange, and the European option contract traded on that index. The source of the option data¹⁴ is the London International Financial Futures and Options Exchange (LIFFE). The data consists of daily closing prices on a wide range of different call and put options, covering the period 4-1-1993 till 28-12-2001, for a total of 902 445 observations. Specifically, the dataset contains the date, call/put flag, strike, option price, expiry month, open interest, volume and the daily settlement price of the FTSE100 index futures contract that has the same maturity as the option contract. The daily settlement time of the futures contract is at 4.30pm; option trading ends at 4.30pm as well. As we select short-maturity at-the-money (ATM) calls which are the most liquid instruments, we expect non-synchronicity biases between the futures and option prices to be negligible. Our analysis also requires the daily returns on the index. The closing price of the index is obtained from DataStream. Again, as trading on the London Stock Exchange ends at 4.30pm, non-synchronicity-biases between the record times of the FTSE100 index value and the closing price of the option are expected to be of minor importance.

Various selection criteria have been applied to extract a call series from the original data base. For each day we have first selected the calls that have a maturity of at least -but at the same time closest to-, 20 trading days¹⁵, i.e., one month. We assume a year to consist of 260 trading days. Second, from these selected calls, we have -for each day- selected the call that is closest ATM by minimizing $|\ln(F_t / K_i)|$, where F_t is the index futures price on that day and K is the strike price. We select ATM calls as these are the ones whose premium is most sensitive for changes in the volatility. Hence, we expect them to contain the most valuable information about the stock volatility. This leaves us with a total of 2258 call price observations¹⁶ for which

¹⁴ We are grateful to Joost Driessen from the Finance Department of the University of Amsterdam for providing us with the option data.

¹⁵ European FTSE100 index option contracts with expiry months March, June, September, December and additional months such that the 3 nearest calendar months are always available for trading, are traded on any given trading day.

¹⁶ Data on 20 trading days are missing in the original database. Furthermore, the data for 1-10-1997 is incomplete and the data for 28-5-1998 contains obvious errors, such that we have discarded these dates.

the maturity ranges from 20 to 44 trading days. The Black-Scholes implied variances are computed using equations (26) and (27), such that estimates of the average dividend yields are not needed. For the interest rates we have performed linear interpolation between the continuously compounded 1-month and 2-month LIBOR rates each day, and taken the rate with term equal to the maturity of the call on that day. These LIBOR rates were taken from DataStream.

Figure 5 displays the squared daily FTSE100 index returns¹⁷ (multiplied by 260) and the Black-Scholes implied variances of the filtered call series. Two things are prominent from this picture. First, there appears to be a clear increase in the level of the volatility after approximately observation 1200 (6-10-1997). The effects of the Asian crisis on the European and global markets that became really noticeable in the fall of 1997, are clearly visible. Furthermore, the combined impact of the crisis in the Soviet Union and the continuing Asian crises, the insecurity about the consequences of the European Monetary Union and the euro, and the insecurity about Clinton's retreatment due to the Monica Lewinsky affair, that all started approximately in the fall of 1998, is clearly visible (after observation 1400). Furthermore, the influence of September 11, 2001 and the subsequent war on terrorism is very obvious. All these events have increased the turbulence on financial markets in the last 5,5 years. Second, the volatility-of-the-volatility seems to have increased during this period in comparison with the period till observation 1100.

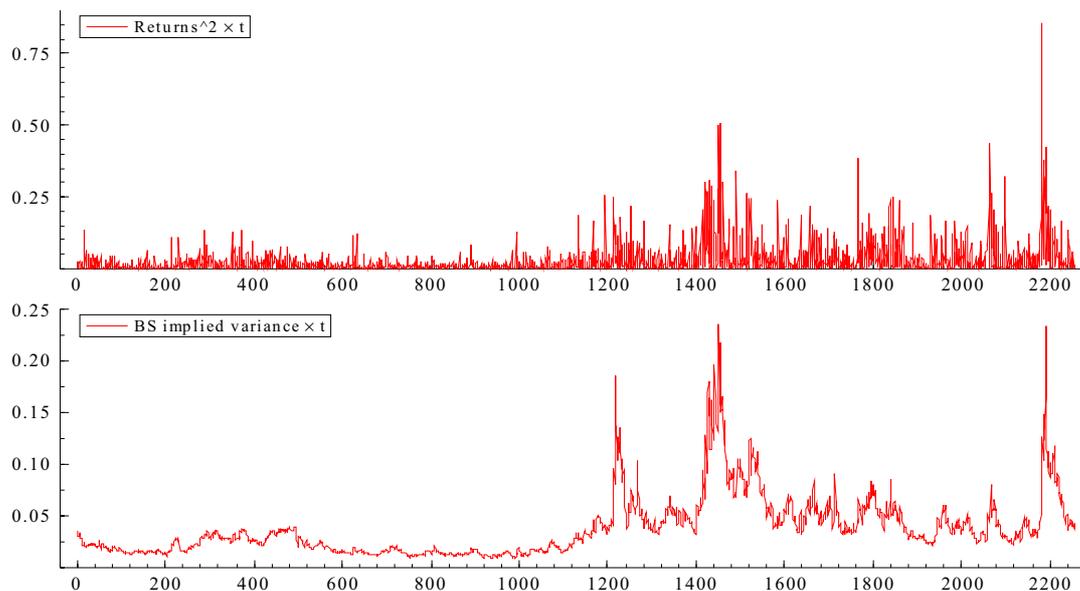


Figure 5: The squared daily FTSE100 returns (multiplied by 260) and the BS implied variances for the period 4-1-1993 till 28-12-2001

¹⁷ The returns have first been "pre-whitened", as follows. There appeared to be significant autocorrelation in the daily returns, believed to be caused by non-synchronous trading effects associated with the individual stocks that constitute together the FTSE100 index; see e.g. Campbell, Lo and MacKinlay (1997). As our model implies non-autocorrelated returns, we have "removed" the autocorrelation by estimating an AR(2) model. The residuals of this regression are taken to be the pre-whitened returns, which are serially uncorrelated.

5.2 The model: 1-factor Ornstein-Uhlenbeck

In the empirical analysis below we study the simplest special case of the affine SV option pricing model in which there is one latent factor that follows an Ornstein-Uhlenbeck (OU) process. In this case $n = 1$, $\alpha_1 = 1$, $\beta_1 = 0$. The basic equations from the continuous-time model become

$$dS_t = \mu S_t dt + \sigma_t S_t dW_{S,t} \quad (\text{P}) \quad (79)$$

$$d\sigma_t^2 = k(\theta - \sigma_t^2)dt + \sigma dW_{\sigma,t} \quad (\text{P})$$

$$d\sigma_t^2 = \tilde{k}(\tilde{\theta} - \sigma_t^2)dt + \sigma d\tilde{W}_{\sigma,t} \quad (\text{Q})$$

$$\tilde{k} = k, \quad \tilde{\theta} = \theta - \frac{\sigma\gamma}{k}$$

$$C_t = \mathbb{E}_{\mathbb{Q}}[BS(S_t, K, \tau, \bar{r}, \bar{q}, \bar{\sigma}^2) | \mathcal{F}_t],$$

where we have further imposed the restrictions $\delta_0 = 0$, $\delta = 1$ such that $\sigma_t^2 = x_t$. For this Gaussian SV model, we can explicitly solve for the functions $A(\tau)$ and $B(\tau)$ that obey the system of ODEs (51)¹⁸. The state space model (78) reduces then to

$$\frac{1}{\Delta t} (r_{t+\Delta t} - \hat{\mu}\Delta t)^2 = \theta + x_t^* + \omega_{t+\Delta t} \quad (80)$$

$$\begin{aligned} \sigma_{implied,t}^2 = \mu_v + \theta + \gamma \frac{\sigma}{k} \left(\frac{1 - \exp(-k\tau)}{k\tau} - 1 \right) + \frac{\sigma^2}{2k^2} \left(1 - 2 \left[\frac{1 - \exp(-k\tau)}{k\tau} \right] + \frac{1 - \exp(-2k\tau)}{2k\tau} \right) \\ + \frac{1 - \exp(-k\tau)}{k\tau} x_t^* + \varepsilon_t \end{aligned} \quad (81)$$

$$x_{t+\Delta t}^* = \exp[-k\Delta t]x_t^* + u_{t+\Delta t}. \quad (82)$$

The error terms satisfy

$$\omega_{t+\Delta t} \sim (0, \sigma_\omega^2); \quad \varepsilon_t \sim (0, \sigma_\varepsilon^2); \quad u_{t+\Delta t} \sim (0, Q), \quad (83)$$

and are both serially and mutually uncorrelated. Also, $\sigma_\omega^2 = \sigma^2/k + 2\theta^2$ and $Q = \sigma^2(1 - \exp(-2k\Delta t))/(2k)$.

Notice that this special case implies the following. The market price of volatility risk is given by the constant γ . As for the general model, the series $\{\sigma_t^2\}$ represents the per annum conditional variance series of the stock returns. The unconditional variance of the stock returns equals $\mathbb{E}[\sigma_t^2] = \mathbb{E}[x_t] = \theta$ per annum, such that the unconditional stock return volatility equals $\sqrt{\theta}$. From the properties of the OU process, the invariant distribution of the variance equals $\sigma_t^2 \sim \mathcal{N}(\theta, \sigma^2/2k)$. Hence, the volatility-of-the-volatility equals $\sqrt{(\sigma^2/2k)}$. Furthermore, the persistence in the daily variance is measured by $\exp[-k\Delta t]$ where $\Delta t = 1/260$. The speed of adjustment towards the mean is given by k .

¹⁸ The calculations are available on request from the first author.

5.3 Estimates based on observations 1200 – 2258

In this section we discuss three sets of parameter estimates for the 1-factor OU special case. These sets are obtained from estimating the model using three alternative sets of data: just return data, just option data, and both sources of data jointly. The estimates are based on observations for 6 October 1997 until 28 December 2001, i.e., observations 1200 – 2258.

For observations 1 - 1200 (subsample 1), an estimate of the unconditional stock volatility based on averaging the squared returns is 11.4%, whereas for observations 1200 – 2258 (subsample 2) the estimate equals 20.3%. Hence, θ has more than tripled in the second sample. Also, as mentioned before, the volatility-of-the-volatility seems to be much larger for the second subsample. Recall that the volatility-of-the-volatility is governed by the parameters σ and k , which we expect to differ over both subsamples. Given these differences between both subsamples, we choose to estimate the model over the second subsample only, such that possible misspecification associated with different θ 's, σ 's and k 's is largely circumvented.

The model is estimated using three types of data. First, the model is estimated using only the return data, such that the model essentially reduces to equations (80) and (82). Second, the model is analyzed using only the option data meaning that the model effectively reduces to equations (81) and (82). Notice that in this case the parameters μ_v and θ cannot separately be identified: only their sum can be estimated. Third, the model is estimated relying on both the return and option data¹⁹. In this case all equations are incorporated.

To allow a comparison, we also estimate a Gaussian GARCH(1,1) model²⁰ for the daily stock returns in which the conditional variance follows the process $\sigma_{t+\Delta t}^2 = \varphi_0 + \varphi_1(r_t - \mu\Delta t)^2 + \varphi_2\sigma_t^2$. We find

$$\sigma_{t+\Delta t}^2 = \underset{(3.18E - 06)}{9.82E - 06} + \underset{(0.0192)}{0.08584} (r_t - \mu\Delta t)^2 + \underset{(0.0348)}{0.8517} \sigma_t^2 \quad (84)$$

with standard errors in parentheses. For this model the persistence in the daily variance is measured by $\varphi_1 + \varphi_2$, whereas the unconditional stock return volatility per annum may be approximated by $\sqrt{\varphi_0 / [(1 - \varphi_1 - \varphi_2)\Delta t]}$.

Table 1 provides the parameter estimates and their robust White (1982) QML standard errors²¹ (in parentheses) associated with the 1-factor OU special case depending on what data is analyzed, together with some other quantities of interest. Note that obviously, not every type of data allows all parameters to be estimated. The fourth column shows some comparable quantities for the estimated GARCH(1,1) model (84).

¹⁹ In the estimation procedure we leave σ_ω as a free parameter to be estimated, as we do not expect it to satisfy the restriction $\sigma_\omega^2 = \sigma^2 / k + 2\theta^2$, since it is obtained from a non-exact Euler discretization. In contrast, as the Euler discretization of the SDE for the factors is exact, we restrict Q in the estimation process to equal the formula given above.

²⁰ Notice that like our model, GARCH does not allow for leverage, permitting a more consistent comparison.

²¹ Their computation is based on Hamilton (1994), section 5.8.

Table 1: Estimation results

	<i>Return data</i>	<i>Option data</i>	<i>Both</i>	<i>GARCH(1,1)</i>
k	16.2 (6.92)	7.12 (5.05)	1.53 (1.70)	
σ	0.164 (0.049)	0.211 (0.0561)	0.147 (0.0208)	
γ	-	0.128 (0.541)	0.133 (0.447)	
σ_ε	-	0.00293 (0.00178)	0.00379 (0.00167)	
σ_ω	0.0612 (0.00561)	-	0.0651 (0.00638)	
θ	0.0409 (0.00520)	-	0.0364 (0.0356)	
μ_v	-	-	0.0174 (0.00329)	
$\mu_v + \theta$	-	0.0572 (0.0136)	0.0538	
Persistence	0.939	0.973	0.994	0.938
Vol. of returns	20.2%	-	19.1%	20.2%
Vol-of-vol	2.87%	5.58%	8.42%	1.75%
\sqrt{Q}	0.00983	0.0129	0.00909	
$\sqrt{\sigma^2 / k + 2\theta^2}$	0.0707	-	0.130	

The table reports the QML parameter estimates (boldface) and their robust QML standard errors (in parentheses), together with some other quantities of interest, resulting from estimating the model using three different types of data: just the squared return data (first column); just the BS implied variance data (second column); both the squared return and the BS implied variance data (third column). The fourth column shows some comparable quantities for the estimated Gaussian GARCH(1,1) model (84).

Looking at the table the following things stand out. The persistence in the daily variance ($\exp[-k\Delta t]$) is estimated to be the smallest (i.e., the mean-reversion the quickest) when the model is estimated using the return data only. In that case the daily persistence equals 0.939, which is equivalent to a speed-of-adjustment coefficient of 16.2. For the GARCH model we find a virtually equal persistence of 0.938. This is not that surprising when we realize that both our model equations (80), (82) and the GARCH(1,1) model can essentially be rewritten as ARMA(1,1) models for the squared returns. If one analyzes the model using the option data only, the estimated persistence of 0.973 is higher, such that the mean-reversion is slower. A possible explanation for this finding may be the fact that squared daily returns are such noisy estimators of the variance that know large swings in value, which causes the estimated speed-of-adjustment towards the mean to be larger, and thus the mean-reversion to be quicker, and hence the persistence to be lower. Notice that when both sources of information are used, the estimated persistence is even larger and the estimated speed of adjustment is thus smaller.

Notice that the volatility-of-the-volatility $\sqrt{(\sigma^2 / 2k)}$ is estimated to be largest when both the return and the option data are used. We may obtain a rough estimate of the per annum volatility-of-the-volatility in the GARCH(1,1) model from computing the standard deviation of the constructed daily GARCH variance series after first having multiplied this series by 260. This estimate equals 1.75%. Compare this value to the volatility-of-the volatility estimated from using the return data only, 2.87%.

The market price of volatility risk γ is estimated to be positive and of the same order of magnitude in both cases, although this parameter is clearly insignificant. Both estimates of σ_ε are roughly the same in magnitude. A similar thing holds for σ_ω . Recall that the parameter μ_v essentially captures the approximation error associated with the call formula. At first glance its value does not seem large, although it differs significantly from zero. Recall that $\sqrt{\theta}$ measures the unconditional volatility of returns, previously approximated at 20.3% based on the average of the squared returns over subsample 2. Note how close this value is to the estimates in the table. The magnitude of the standard deviation of the transition equation disturbance $u_{t+\Delta t}$, \sqrt{Q} , is roughly the same for all cases. The final row in the table displays the estimated theoretical value of $\sigma_\omega^2 = \sigma^2 / k + 2\theta^2$, which should be compared to the unrestrictedly estimated σ_ω .

5.4 The volatility series and the in-sample fit of the model estimated with both return- and option data

Let us now investigate the extracted volatility series and the in-sample fit of our model estimated using both the return- and the option data. We concentrate on the smoothed estimates of all quantities of interest. Figure 6 displays the smoothed stock volatility series $\sigma_t = \sqrt{X_{t|T}^* + \theta}$, the (annualized) GARCH volatility series obtained from (84), and the original BS implied volatility data²².

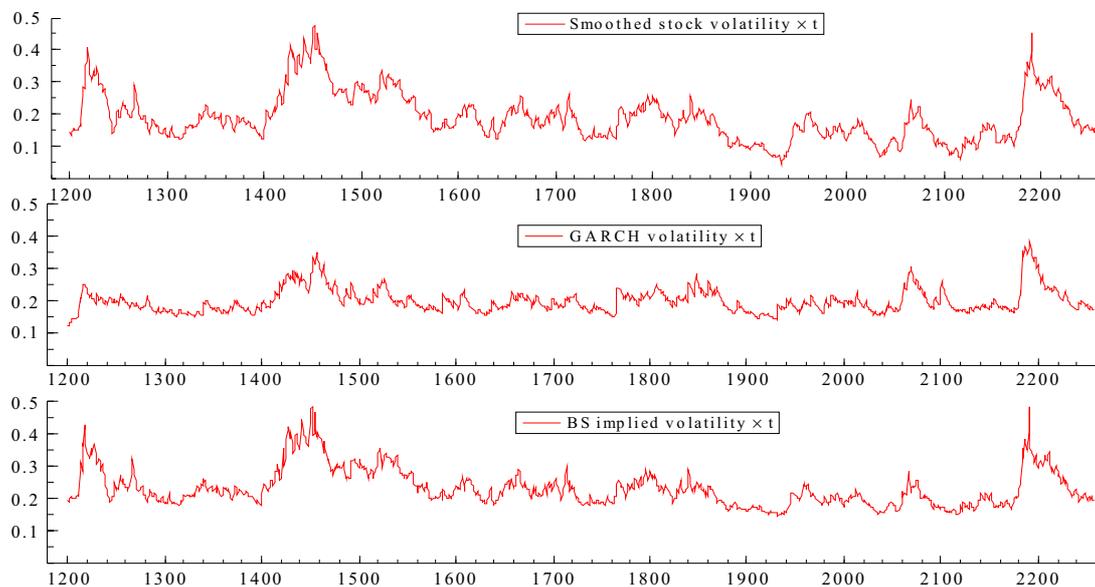


Figure 6: The smoothed stock return volatility series $\sigma_t = (X_{t|T}^* + \theta)^{1/2}$, the GARCH volatility series, and the Black-Scholes implied volatility data.

²² Recall that $\{\sigma_t^2\}$ represents the conditional variance series of the stock returns (per annum). The annualized GARCH volatility series has been obtained from multiplying the daily GARCH volatility series obtained from (84) with $\sqrt{260}$.

Notice the rough analogy in the patterns of the different volatility series, although the GARCH series looks somewhat “smoother” and does not peak as much as both other series. Besides, in contrast to the smoothed volatilities from our model, the GARCH volatilities seem to have some lower “barrier” that is virtually never exceeded. This “smoother” behavior may be caused by the fact that the impact of a time t shock in the conditional mean (i.e., $r_t - \mu\Delta t$) on the time $t + \Delta t$ GARCH conditional variance is given by $\phi_1(r_t - \mu\Delta t)^2$ which is generally small. In contrast, the SV model allows for an autonomous shock $u_{t+\Delta t}$ each period that fully impacts on the SV conditional variance. Therefore, the SV model seems better able to describe quickly changing patterns, which is evidenced by figure 6: the smoothed volatilities track the observed BS implied volatility data better than the GARCH volatilities do. However, to obtain more insight in the differences between the three series, in figure 7 we plot the difference between the BS implied volatility data and the smoothed volatilities (upper graph), the difference between the BS implied volatility data and the GARCH volatilities (middle graph), and the difference between the GARCH- and the smoothed volatilities (lower graph). Notice that the scales of the vertical axis coincide.

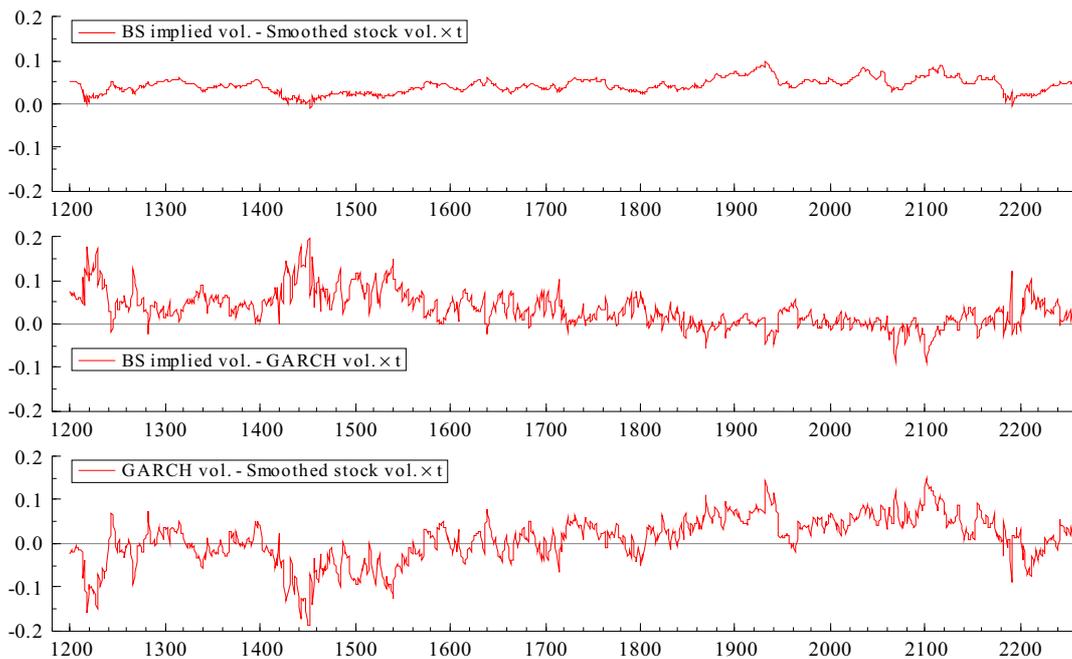


Figure 7: Upper graph: the difference between the BS implied volatility data and the smoothed volatilities $\sigma_t = (X_{t|T}^* + \theta)^{1/2}$. Middle graph: the difference between the BS implied volatility data and the GARCH volatilities. Lower graph: the difference between the GARCH- and the smoothed volatilities.

We note the following. The observed Black-Scholes implied volatilities are generally larger than the smoothed volatilities. Hence, it seems that the market expectation of the volatility is generally higher than the “true” volatility. Also, the BS implied volatilities are generally larger than the GARCH volatilities, although in the second

part of the sample this is less so (but remember that “virtual lower barrier” of the GARCH volatilities). The difference between the GARCH- and the smoothed volatilities fluctuates much more around zero. A reason for these findings may be the fact that both the GARCH- and our SV model allow fat-tailedness in stock returns, whereas the conventional BS model with which we computed the BS implied volatilities does not.

Figure 8 provides insight in how well the model fits the original data. The upper graph displays the squared return data and its inherent signal, the smoothed stock return variance $\sigma_t^2 = X_{t|T}^* + \theta$. The lower graph shows the observed BS implied variance data and the fitted (smoothed) BS implied variances computed from equation (81).

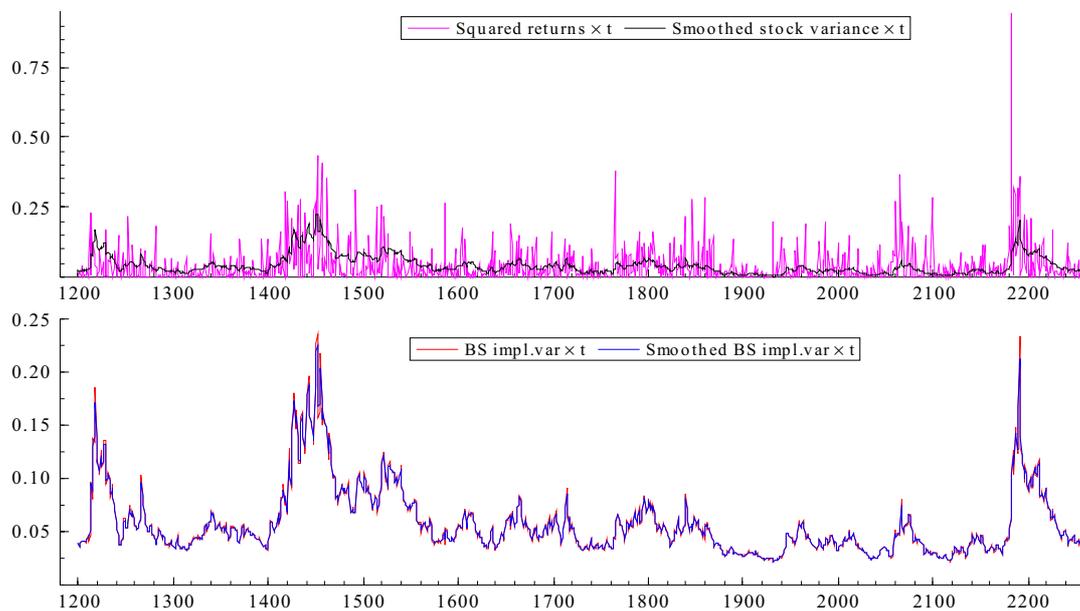


Figure 8: Upper plot: the squared return data and its inherent signal, the smoothed stock return variance $\sigma_t^2 = X_{t|T}^* + \theta$. Lower plot: the BS implied variance data and the fitted BS implied variances.

Notice how well the model seems to fit the data. The two series in the lower picture are virtually indistinguishable. In particular, OLS regression of the observed BS implied variances on the smoothed BS implied variances yields a constant term of -0.00072 , a slope of 1.01 and an R-squared of 0.99634 .

5.5 Diagnostic checking

In this section diagnostic checking is considered. Given the parameter estimates and the Kalman filter output the standardized prediction errors or innovations²³ can be constructed. If the model is well-specified these errors ought to be a mean zero unit

²³ In terms of the general state space model, the standardized prediction errors are defined as $(\mathbf{H}_t' \mathbf{P}_{t|t-1} \mathbf{H}_t + \mathbf{R})^{-\frac{1}{2}} (\mathbf{y}_t - \hat{\mathbb{E}}[\mathbf{y}_t | \mathcal{Y}_{t-1}]) = (\mathbf{H}_t' \mathbf{P}_{t|t-1} \mathbf{H}_t + \mathbf{R})^{-\frac{1}{2}} (\mathbf{y}_t - \mathbf{a}_t - \mathbf{H}_t' \hat{\boldsymbol{\xi}}_{t|t-1})$.

variance uncorrelated series. Figure 9 displays the errors associated with the squared return equation (80) in the upper graph, and the errors associated with the BS implied variance equation (81) in the lower graph. The means of the series equal 0.000, respectively 0.003, whereas the standard deviations equal 1.005, respectively 0.996. Both series are clearly heteroskedastic.

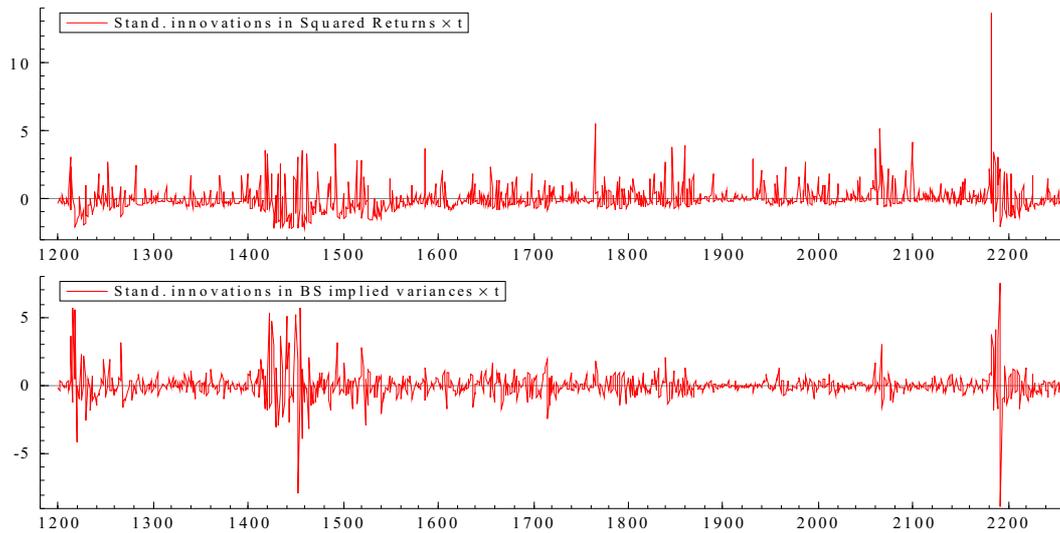


Figure 9: The standardized prediction errors belonging to the squared return equation (80) (upper graph) and the BS implied variance equation (81) (lower graph).

Furthermore, as evidenced by table 2, both series are significantly autocorrelated. For example, the p-values (Prob) associated with the Ljung-Box Q-statistics (Q-stat) for testing the null hypothesis of there being no autocorrelation up to order 5 equals 0.000 for both series. The column headed AC contains the first to fifth order autocorrelation coefficients.

Table 2: Autocorrelations in the standardized prediction error series and associated statistics

<i>Statistics for the standardized innovations in the squared return equation:</i>				<i>Statistics for the standardized innovations in the BS implied variance equation:</i>			
Order	AC	Q-Stat	Prob	Order	AC	Q-Stat	Prob
1	0.081	7.0415	0.008	1	0.034	1.2226	0.269
2	0.053	10.001	0.007	2	-0.147	24.150	0.000
3	0.130	27.877	0.000	3	-0.106	36.129	0.000
4	0.109	40.613	0.000	4	0.038	37.704	0.000
5	0.048	43.019	0.000	5	-0.004	37.718	0.000

The table reports the autocorrelation coefficients (AC) up to order 5 and the Ljung-Box Q-statistics (Q-stat) for testing the null of no autocorrelation up to a certain order with associated p-values (Prob). The left (resp. right) part of the table deals with the standardized innovations in the squared return (resp. BS implied variance) equation.

The fact that both series are significantly autocorrelated is evidence of misspecification, and may be interpreted as a preliminary indication of neglected dynamics in this 1-factor Ornstein-Uhlenbeck model.

Besides the standardized prediction errors, the so-called smoothed disturbances or auxiliary residuals yield additional information with respect to model misspecification. They represent the best estimates of the disturbances in the state space model given the data²⁴. Figure 10 displays the auxiliary residuals belonging to the measurement equation.

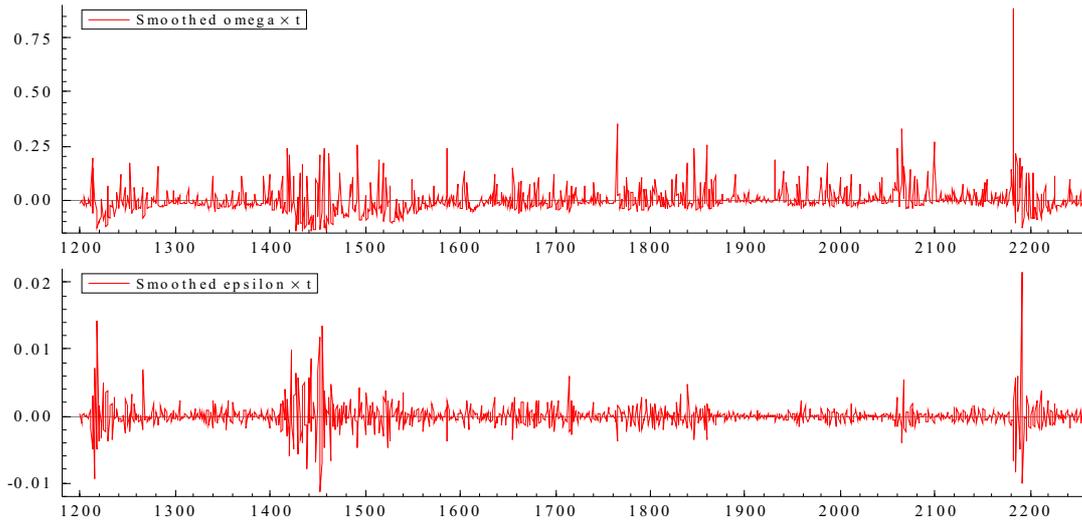


Figure 10: The auxiliary residuals of the measurement equation: the smoothed $\{\omega_{t+\Delta t}\}$ series associated with equation (80) (upper plot), and the smoothed $\{\varepsilon_t\}$ series associated with eq. (81) (lower plot).

Notice that there is evidence of conditional heteroskedasticity. The asymmetric pattern in the smoothed $\{\omega_{t+\Delta t}\}$ series is a consequence of the course of the squared return series over time; recall figure 5. When an ARMA(1,1) model is fitted to the squared returns a similar picture results for the residuals.

Figure 11 displays the smoothed state $x_{t|T}^*$ together with the smoothed state disturbance $\mathbb{E}[u_{t+\Delta t} | \mathcal{Y}_T]$. Inspecting figure 11 it appears that when the level of the state is high, the state disturbance is large (in absolute value). Hence, the variance of the state disturbance seems to increase if the level of the state is high. Recalling that the state governs the conditional stock return variance, this is evidence of *level-dependent volatility*.

²⁴ In terms of the general state space model, the smoothed disturbances (or auxiliary residuals) of the measurement equation are defined as $\mathbb{E}[\mathbf{w}_t | \mathcal{Y}_T] = \mathbf{y}_t - \mathbf{a}_t - \mathbf{H}_t' \hat{\boldsymbol{\xi}}_{t|T}$. For the transition equation they equal $\mathbb{E}[\mathbf{v}_{t+1} | \mathcal{Y}_T] = \hat{\boldsymbol{\xi}}_{t+1|T} - \mathbf{F} \hat{\boldsymbol{\xi}}_{t|T}$. As Durbin and Koopman (2001) point out they are autocorrelated but can be useful for detecting outliers and structural breaks as they represent estimators of the error term.

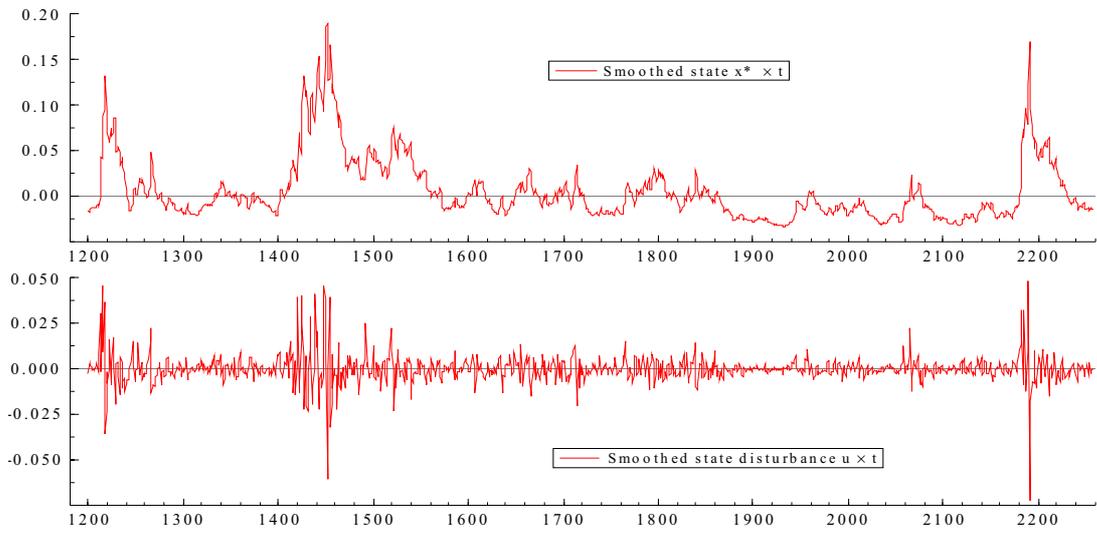


Figure 11: The smoothed state $X_{t|T}^*$ (upper graph) and the smoothed state disturbance $\mathbb{E}[u_{t+\Delta t} | \mathcal{Y}_T]$.

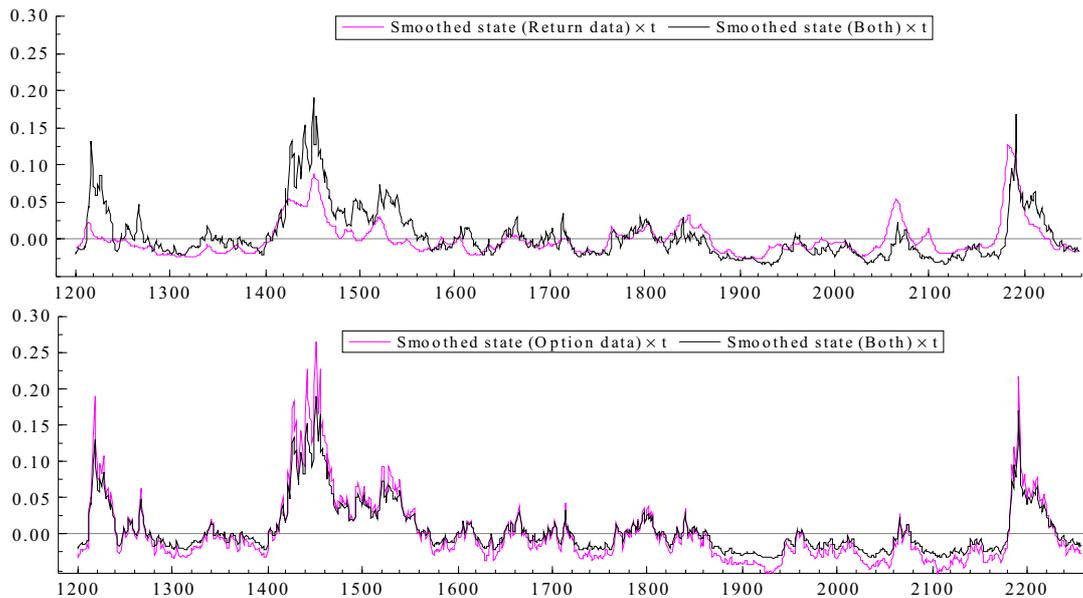


Figure 12: The upper plot displays the smoothed state obtained from estimating the model using only the return data, together with the smoothed state obtained from using both the return and option data. The lower plot shows the smoothed state obtained from using only the option data, together with the smoothed state obtained when using both sources of data.

Finally, it is instructive to compare the smoothed states of the model estimated with the three different types of data. The upper plot in figure 12 displays the obtained smoothed state when only the return data is used for estimation, together with the smoothed state when both the return and the option data are used. Notice that they seem to deviate pretty much from time to time. In the lower graph, the smoothed

state when only the option data is used is plotted, together with the state obtained from using both sources of data. Notice that these latter smoothed states seem to track each other much better. It seems that the information in the option data dominates the estimation results when both the return- and the option data are used. This may be attributed to the fact that squared daily returns are such noisy estimators of the variance: they contain less precise information and therefore less pronounced structure.

6. Summary and Directions for Future Research

In this paper we introduce the class of multi-factor affine SV option pricing models. This class is more flexible than the more conventional models, although it does not allow for the leverage effect yet. It is assumed that the variance of the stock returns is driven by an affine function of an arbitrary number of latent factors, which follow stationary mean-reverting Markov diffusions. The inspiration for proposing this class was found in the literature dealing with the term structure of interest rates. A major advantage of this class is that it allows to investigate if the volatility is driven by more than one risk factor in practice. We derive a call pricing formula for this class of SV option pricing models.

Next, we propose a method to estimate the parameters of such models. We write the model in linear state space form such that it can be analyzed by the Kalman filter and smoother. An important step in this process is to linearize the Black-Scholes pricing function, viewed as a function of the exponent of the integrated variance, around a suitably chosen value. This linearization subsequently ensures that we end up with a partly explicit analytical expression for the call premium. We provide some promising results that a linearization may not be a too bad approximation. The benefit of the state space approach is that it allows to simultaneously exploit the information in both the time series of option prices and the time series of the underlying stock. Moreover, it readily delivers the filtered or smoothed volatility series. We argue that our method may be considered an attractive alternative to EMM.

We use daily data on the FTSE100 index to illustrate the method. We provide estimation results for a simple special case of the affine class, in which there is one factor driving the volatility that follows an Ornstein-Uhlenbeck process. The model is estimated using only return data, using just option data, and finally using both sources of data. The extracted smoothed volatility series is contrasted to both an estimated GARCH volatility series and to observed Black-Scholes implied volatility data. The graphs we supply indicate that the method works well.

Diagnostic checks reveal a number of interesting issues that will be explored in our future research. First, we find evidence of neglected dynamics in the empirical analysis of the 1-factor OU model. This motivates the empirical investigation of multi-factor models which may be particularly relevant when considering a panel of option price series with different maturities and moneyness. Second, the results indicate the existence of level-dependent volatility. This motivates the investigation of stock variance driving processes that explicitly model this level-dependent volatility, like the Cox-Ingersoll-Ross process for example. In turn, this calls for the *Extended Kalman Filter* which can handle conditional distributions. Third, we find that the option data dominates the estimation results when the model is estimated using

both return and option data. We argue that a reason for this finding may be the fact that squared daily returns are such noisy estimators of the variance. This motivates the use of high-frequency intraday return data and realized volatility measures in our analysis. As these measures are far less noisy, we expect the information in the realized volatility and option data to be more equally weighted in obtaining the parameter estimates, than in case of using squared daily returns. Together with an extension towards incorporating the leverage effect and towards considering a panel of options, this will be investigated in our future research.

Appendix

A. Our model is arbitrage-free: a proof

In section 2.2, we implicitly proposed a possible choice for the non-unique \mathbb{Q} by simultaneously assuming a particular form for the market price of stock risk and the market price of volatility risk, and claimed that the resulting model is arbitrage-free. Here we prove this claim by showing that the prices of the tradable assets expressed in terms of the money market account are \mathbb{Q} -martingales, as they should be according to the Fundamental Theorem of Asset Pricing.

First, and obviously, the relative price of the money market account equals 1 for all t under all measures, such that it is a martingale irrespective of the specific measure. Second, note that the other tradable asset is the dividend earning stock. An important notice is the fact that the process $\{S_t; t \geq 0\}$ itself is not a tradable asset, because it does not account for dividend payments. Consider the following portfolio strategy. Suppose at time 0 we buy a_0 (some positive number) units of the stock, and that all the dividends we receive are immediately reinvested to buy additional units of the stock. Doing so, at time t we have

$$a_t \equiv a_0 \exp\left(\int_0^t q_u du\right) \quad (\text{A1})$$

units of the stock in our so-called reinvestment portfolio, which we denote by S^r . Notice that this reinvestment portfolio is indeed a tradable asset. The value of this asset at time t equals $S_t^r = a_t S_t$. Notice that $S_t^r = f(t, S_t)$. Invoking Itô's lemma yields

$$dS_t^r = (\mu + q_t)S_t^r dt + \sigma_t S_t^r dW_{S,t}. \quad (\mathbb{P}) \quad (\text{A2})$$

Consider now the relative price Z_t^r of the reinvestment portfolio, defined by

$$Z_t^r \equiv B_t^{-1} S_t^r. \quad (\text{A3})$$

Using Itô, it can be shown that it follows the process

$$dZ_t^r = (\mu + q_t - r_t)Z_t^r dt + \sigma_t Z_t^r dW_{S,t} \quad (\mathbb{P}) \quad (\text{A4})$$

under \mathbb{P} , whereas under our choice of \mathbb{Q} , for which

$$d\tilde{W}_{S,t} = dW_{S,t} + \gamma_{S,t} dt; \quad \gamma_{S,t} \equiv \frac{\mu + q_t - r_t}{\sigma_t}, \quad (\text{A5})$$

it follows

$$dZ_t^r = \sigma_t Z_t^r d\tilde{W}_{S,t}. \quad (\text{Q}) \quad (\text{A6})$$

(Notice that the specific form of the market price of volatility risk does not play a role in the argument.) This SDE does not contain a drift such that $\{Z_t^r; t \geq 0\}$ is a martingale process under \mathbb{Q} . Therefore our model is indeed arbitrage-free. This completes the proof.

B. Proof of the series $\{\omega_{t+\Delta t}\}$ being white noise

Below we prove that the series $\{\omega_{t+\Delta t}\}$ is a white noise series. First, notice that $\omega_{t+\Delta t}$ can be written as

$$\omega_{t+\Delta t} \equiv \frac{1}{\Delta t} (r_{t+\Delta t} - \mu \Delta t)^2 - \sigma_t^2 = \sigma_t^2 \left(\frac{\eta_{t+\Delta t}^2}{\Delta t} - 1 \right). \quad (\text{B1})$$

It is not difficult to see that $\{\omega_{t+\Delta t}\}$ is a covariance-stationary series. It has mean zero, $\mathbb{E}[\omega_{t+\Delta t}] = 0$. For the variance we find

$$\text{var}[\omega_{t+\Delta t}] = \mathbb{E}[\sigma_t^4 \left(\frac{\eta_{t+\Delta t}^2}{\Delta t} - 1 \right)^2] = \mathbb{E}[\sigma_t^4] \mathbb{E}[\left(\frac{\eta_{t+\Delta t}^2}{\Delta t} - 1 \right)^2], \quad (\text{B2})$$

where we made use of the independence of σ_t and $\eta_{t+\Delta t}$. Now,

$$\mathbb{E}[\left(\frac{\eta_{t+\Delta t}^2}{\Delta t} - 1 \right)^2] = \mathbb{E}[\frac{\eta_{t+\Delta t}^4}{\Delta^2 t} - 2 \frac{\eta_{t+\Delta t}^2}{\Delta t} + 1] = \frac{3\Delta^2 t}{\Delta^2 t} - 2 \frac{\Delta t}{\Delta t} + 1 = 2, \quad (\text{B3})$$

since $\mathbb{E}[\eta_{t+\Delta t}^4] = 3\Delta^2 t$. Also,

$$\mathbb{E}[\sigma_t^4] = \mathbb{E}[(\delta_0 + \boldsymbol{\delta}' \mathbf{x}_t)^2] = \delta_0^2 + 2\delta_0 \boldsymbol{\delta}' \boldsymbol{\theta} + \boldsymbol{\delta}' (\text{var}[\mathbf{x}_t] + \boldsymbol{\theta} \boldsymbol{\theta}') \boldsymbol{\delta} \quad (\text{B4})$$

In another paper²⁵ we performed tedious algebra to show that the unconditional variance of the factors equals

$$\text{var}[\mathbf{x}_t] = \mathbf{J} \odot \boldsymbol{\Sigma} \mathbf{M}_d \boldsymbol{\Sigma}', \quad (\text{B5})$$

in which \mathbf{J} is an $(n \times n)$ matrix having its ij -th element equal to $[\mathbf{J}]_{ij} = 1/(k_i + k_j)$, and where \odot represents the Hadamard product; i.e., element-by-element multiplication. Hence, the variance of $\omega_{t+\Delta t}$ equals

$$\text{var}[\omega_{t+\Delta t}] = 2\delta_0^2 + 4\delta_0 \boldsymbol{\delta}' \boldsymbol{\theta} + 2\boldsymbol{\delta}' (\mathbf{J} \odot \boldsymbol{\Sigma} \mathbf{M}_d \boldsymbol{\Sigma}' + \boldsymbol{\theta} \boldsymbol{\theta}') \boldsymbol{\delta} \equiv \sigma_\omega^2. \quad (\text{B6})$$

Note that the series $\{\omega_{t+\Delta t}\}$ is homoskedastic as it should be because of its covariance-stationarity. Consider next the autocovariance of order $p = 1, 2, \dots$,

$$\begin{aligned} \text{cov}[\omega_{t+\Delta t}, \omega_{t-(p-1)\Delta t}] &= \mathbb{E}[\sigma_t^2 \left(\frac{\eta_{t+\Delta t}^2}{\Delta t} - 1 \right) \sigma_{t-\rho\Delta t}^2 \left(\frac{\eta_{t-(p-1)\Delta t}^2}{\Delta t} - 1 \right)] \\ &= \mathbb{E}[\sigma_t^2 \sigma_{t-\rho\Delta t}^2] \mathbb{E}[\left(\frac{\eta_{t+\Delta t}^2}{\Delta t} - 1 \right) \left(\frac{\eta_{t-(p-1)\Delta t}^2}{\Delta t} - 1 \right)] = 0, \end{aligned} \quad (\text{B7})$$

²⁵ These calculations are available on request from Antoine P.C. van der Ploeg.

where we made use of the fact that the volatility process is independent of $\{\omega_{t+\Delta t}\}$, and that $\eta_{t+\Delta t} \sim \text{i.i.d. } \mathcal{N}(0, \Delta t)$. Hence the series is serially uncorrelated. To conclude, we find that the error term series $\{\omega_{t+\Delta t}\}$ is indeed a white noise series.

C. The statistical properties of the error series $\{\mathbf{u}_{t+\Delta t}\}$

Below we investigate the statistical properties of the error term series $\{\mathbf{u}_{2\Delta t}, \dots, \mathbf{u}_{(T+1)\Delta t}\}$ belonging to the transition equation. Recall its definition in equation (74). For notational simplicity in the computations, we let t run from $t = 2\Delta t, 3\Delta t, \dots, (T+1)\Delta t$. As the expected value of Itô stochastic integrals equals zero, the mean of \mathbf{u}_t equals zero,

$$\mathbb{E}[\mathbf{u}_t] = \mathbf{0}; \quad t = 2\Delta t, \dots, (T+1)\Delta t \quad (\text{C1})$$

To obtain its variance we make use of the following definitions,

$$\mathbf{C}_d(u, t) \equiv \exp[-\mathbf{K}_d(t-u)] = \text{diag}[c_1(u, t), \dots, c_n(u, t)]. \quad (\text{C2})$$

$$\mathbf{c}(u, t) \equiv [c_1(u, t), \dots, c_n(u, t)]'$$

$$c_i(u, t) \equiv \exp[-k_i(t-u)], \quad i = 1, \dots, n.$$

The variance of \mathbf{u}_t is then computed as follows,

$$\begin{aligned} \text{var}[\mathbf{u}_t] &= \text{var} \left[\int_{t-\Delta t}^t \mathbf{C}_d(u, t) \boldsymbol{\Sigma} \boldsymbol{\Lambda}_u d\mathbf{W}_{x,u} \right] = \int_{t-\Delta t}^t \mathbf{C}_d(u, t) \boldsymbol{\Sigma} \mathbb{E}[\boldsymbol{\Lambda}_u^2] \boldsymbol{\Sigma}' \mathbf{C}_d(u, t) du \\ &= \int_{t-\Delta t}^t \mathbf{C}_d(u, t) \boldsymbol{\Sigma} \mathbf{M}_d \boldsymbol{\Sigma}' \mathbf{C}_d(u, t) du, \end{aligned} \quad (\text{C3})$$

in which, after remembering that $\mathbb{E}[\mathbf{x}_t] = \boldsymbol{\theta}$, the $(n \times n)$ diagonal matrix \mathbf{M}_d is defined as $\mathbf{M}_d \equiv \mathbb{E}[\boldsymbol{\Lambda}_u^2] = \text{diag}[\alpha_1 + \boldsymbol{\beta}_1' \boldsymbol{\theta}, \dots, \alpha_n + \boldsymbol{\beta}_n' \boldsymbol{\theta}]$. Before proceeding, it is not difficult to show that the following result holds. For a general $(n \times n)$ diagonal matrix $(n \times n) \mathbf{D}_d = \text{diag}[d_1, \dots, d_n]$, and vector $(n \times 1) \mathbf{d} = [d_1, \dots, d_n]'$ and general matrix $(n \times n) \mathbf{A}$, we have that $\mathbf{D}_d \mathbf{A} \mathbf{D}_d = \mathbf{d} \mathbf{d}' \odot \mathbf{A}$. Making use of this result we subsequently obtain

$$\text{var}[\mathbf{u}_t] = \int_{t-\Delta t}^t [\mathbf{c}(u, t) \mathbf{c}(u, t)' \odot \boldsymbol{\Sigma} \mathbf{M}_d \boldsymbol{\Sigma}'] du = \left(\int_{t-\Delta t}^t \mathbf{c}(u, t) \mathbf{c}(u, t)' du \right) \odot \boldsymbol{\Sigma} \mathbf{M}_d \boldsymbol{\Sigma}' = \mathbf{G} \odot \boldsymbol{\Sigma} \mathbf{M}_d \boldsymbol{\Sigma}' \quad (\text{C4})$$

where the $(n \times n)$ matrix \mathbf{G} has its ij -th element $[\mathbf{G}]_{ij}$ equal to

$$[\mathbf{G}]_{ij} = \int_{t-\Delta t}^t c_i(u, t) c_j(u, t) du = \int_{t-\Delta t}^t \exp[-(k_i + k_j)(t-u)] du = \frac{1 - \exp[-(k_i + k_j)\Delta t]}{k_i + k_j}. \quad (\text{C5})$$

Notice that the variance is time invariant such that the series $\{\mathbf{u}_t\}$ is homoskedastic. Next, consider the autocovariance of order p ($p = 1, 2, \dots$), given by

$$\text{cov}[\mathbf{u}_t, \mathbf{u}_{t-\rho\Delta t}] = \mathbb{E}[\mathbf{u}_t \mathbf{u}_{t-\rho\Delta t}'] ; \quad t = 2\Delta t, \dots, (T+1)\Delta t \quad (\text{C6})$$

This expectation can be tackled by conditioning on $\mathcal{F}_{t-\rho\Delta t}$ to get

$$\text{cov}[\mathbf{u}_t, \mathbf{u}_{t-\rho\Delta t}] = \mathbb{E}[\mathbb{E}(\mathbf{u}_t \mathbf{u}_{t-\rho\Delta t}' | \mathcal{F}_{t-\rho\Delta t})] = \mathbb{E}[\mathbb{E}(\mathbf{u}_t | \mathcal{F}_{t-\rho\Delta t}) \mathbf{u}_{t-\rho\Delta t}'] . \quad (\text{C7})$$

Can we proceed any further? Yes, we can. First, consider the following expectation

$$\begin{aligned} & \mathbb{E} \left[\int_0^t \mathbf{C}_d(u, t) \boldsymbol{\Sigma} \boldsymbol{\Lambda}_u d\mathbf{W}_{x,u} | \mathcal{F}_{t-\rho\Delta t} \right] \quad (\text{C8}) \\ &= \mathbb{E} \left[\int_0^{t-\rho\Delta t} \mathbf{C}_d(u, t) \boldsymbol{\Sigma} \boldsymbol{\Lambda}_u d\mathbf{W}_{x,u} + \int_{t-\rho\Delta t}^{t-\Delta t} \mathbf{C}_d(u, t) \boldsymbol{\Sigma} \boldsymbol{\Lambda}_u d\mathbf{W}_{x,u} + \int_{t-\Delta t}^t \mathbf{C}_d(u, t) \boldsymbol{\Sigma} \boldsymbol{\Lambda}_u d\mathbf{W}_{x,u} | \mathcal{F}_{t-\rho\Delta t} \right] \\ &= \int_0^{t-\rho\Delta t} \mathbf{C}_d(u, t) \boldsymbol{\Sigma} \boldsymbol{\Lambda}_u d\mathbf{W}_{x,u} + \mathbb{E} \left[\int_{t-\rho\Delta t}^{t-\Delta t} \mathbf{C}_d(u, t) \boldsymbol{\Sigma} \boldsymbol{\Lambda}_u d\mathbf{W}_{x,u} | \mathcal{F}_{t-\rho\Delta t} \right] + \mathbb{E}[\mathbf{u}_t | \mathcal{F}_{t-\rho\Delta t}] \end{aligned}$$

Remembering that Itô integrals are martingales with respect to the natural filtration of the Brownian motion driving the integral, it must be the case that

$$\mathbb{E} \left[\int_{t-\rho\Delta t}^{t-\Delta t} \mathbf{C}_d(u, t) \boldsymbol{\Sigma} \boldsymbol{\Lambda}_u d\mathbf{W}_{x,u} | \mathcal{F}_{t-\rho\Delta t} \right] = \mathbf{0}, \quad \mathbb{E}[\mathbf{u}_t | \mathcal{F}_{t-\rho\Delta t}] = \mathbf{0} . \quad (\text{C9})$$

But this implies that

$$\text{cov}[\mathbf{u}_t, \mathbf{u}_{t-\rho\Delta t}] = \mathbf{0} ; \quad t = 2\Delta t, \dots, (T+1)\Delta t . \quad (\text{C10})$$

A similar argument can be used to show that $\text{cov}[\mathbf{u}_t, \mathbf{u}_{t+\rho\Delta t}] = \mathbf{0}$. Thus, the error term series $\{\mathbf{u}_t; t = 2\Delta t, \dots, (T+1)\Delta t\}$ is serially uncorrelated.

Recalling the state space model formulation, we should also check if the error term at time t is uncorrelated with lagged values of the state. For this we need to assume that the initial state is uncorrelated with the error term series, as in, e.g., Hamilton (1994): $\mathbb{E}[\mathbf{u}_t, \mathbf{x}_{\Delta t}^*'] = \mathbf{0} \quad \forall t = 2\Delta t, \dots, (T+1)\Delta t$. It is then relatively easy to show that this indeed holds in the present setting. Namely, we have for $p = 1, 2, \dots$,

$$\text{cov}[\mathbf{u}_t, \mathbf{x}_{t-\rho\Delta t}^*] = \mathbb{E}[\mathbb{E}(\mathbf{u}_t \mathbf{x}_{t-\rho\Delta t}^* | \mathcal{F}_{t-\rho\Delta t})] = \mathbb{E}[\mathbb{E}(\mathbf{u}_t | \mathcal{F}_{t-\rho\Delta t}) \mathbf{x}_{t-\rho\Delta t}^*] = \mathbf{0}, \quad (\text{C11})$$

where we made use of the fact that $\mathbb{E}[\mathbf{u}_t | \mathcal{F}_{t-\rho\Delta t}] = \mathbf{0}$ which we just proved.

What remains to be checked is the correlation between the errors of the measurement equation and the errors of the transition equation. These ought to be zero in the state space formulation. And indeed they are, as we will now argue. The error term of the measurement equation is given by $\mathbf{w}_t \equiv (\omega_{t+\Delta t}, \varepsilon_t)'$, whereas the error term of the transition equation is given by $\mathbf{u}_{t+\Delta t}$. By assumption, the series $\{\varepsilon_t\}$ is uncorrelated with any other series. For the covariance between $\omega_{t+\Delta t}$ and $\mathbf{u}_{s+\Delta t}$, $s, t = \Delta t, \dots, (T+1)\Delta t$, we find

$$\text{cov}[\omega_{t+\Delta t}, \mathbf{u}_{S+\Delta t}] = \mathbb{E}[\sigma_t^2 (\frac{\eta_{t+\Delta t}^2}{\Delta t} - 1) \mathbf{u}_{S+\Delta t}] = \mathbb{E}[\sigma_t^2 \mathbf{u}_{S+\Delta t}] \mathbb{E}[\frac{\eta_{t+\Delta t}^2}{\Delta t} - 1] = \mathbf{0}. \quad (\text{C12})$$

The argument is as follows. The error term $\mathbf{u}_{S+\Delta t}$ and the volatility σ_t are driven by the Brownian motions \mathbf{W}_x and not by the Brownian motion driving the stock price, W_S , which determines $\eta_{t+\Delta t}$. Hence, the product $\sigma_t^2 \mathbf{u}_{S+\Delta t}$ is independent of $\eta_{t+\Delta t}$, from which the result follows.

References

- Ball, C.A. and A. Roma, 1994, Stochastic volatility option pricing, *Journal of Financial and Quantitative Analysis* 29, 589-607.
- Bakshi, G., C. Cao and Z. Chen, 1997, Empirical performance of alternative option pricing models, *Journal of Finance* vol. 52, issue 5, 2003 – 2049.
- Bates, D. S., 1996, Testing option pricing models, in: Maddala, G.S., C.R. Rao, eds., *Handbook of Statistics*, Vol. 14 (Elsevier, Amsterdam).
- Bates, D. S., 2000, Post-'87 crash fears in S&P500 futures options, *Journal of Econometrics* 94 (1-2), 181 – 238.
- Baxter, M. and A. Rennie, 1996, *Financial calculus: an introduction to derivative pricing* (Cambridge University Press, Cambridge UK).
- Black, F. and M. Scholes, 1973, The pricing of options and corporate liabilities, *The Journal of Political Economy* 81, 637-54.
- Campbell, J.Y., A.W. Lo and A.C. MacKinlay, 1997, *The econometrics of financial markets* (Princeton University Press, Princeton NJ).
- Chernov, M. and E. Ghysels, 2000, A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of options valuation, *Journal of Financial Economics* 56, 407 – 458.
- Chesney, M. and L. Scott, 1989, Pricing european currency options: A comparison of the modified Black-Scholes model and a random variance model, *Journal of Financial and Quantitative Analysis* 24, 267-84.
- Dai, Q. and K.J. Singleton, 2000, Specification analysis of affine term structure models, *The Journal of Finance*, Vol. LV, no. 5, 1943 – 1978.
- Delbaen, F. and W. Schachermayer, 1994, The fundamental theorem of asset pricing, *Mathematische Annalen* 300, 463-520.
- Duffee, G. and R. H. Stanton, 2001, Estimation of dynamic term structure models, Unpublished paper, Haas School of Business, U.C. Berkeley, California, USA.
- Duffie, D. and R. Kan, 1996, A yield-factor model of interest rates, *Mathematical Finance* 6, 379 – 406.
- Duffie, D., J. Pan and K. Singleton, 2000, Transform analysis and asset pricing for affine jump diffusions, *Econometrica* 68, 1343-1376.
- Duffie, D. and K.J. Singleton, 1993, Simulated moments estimation of Markov models of asset prices, *Econometrica* 61, 929-52.
- Durbin, J. and S.J. Koopman, 2001, *Time Series Analysis by State Space Methods*, (Oxford University Press, Oxford, UK).
- Gallant, A.R., D. Hsieh and G. Tauchen, 1997, Estimation of stochastic volatility models with diagnostics, *Journal of Econometrics* 81, 159-92.
- Gallant, A.R. and G. Tauchen, 1996, Which moments to match?, *Econometric Theory* 12, 657 – 681.
- Ghysels, E., Harvey, A.C. and E. Renault, 1996, Stochastic volatility, in: *Statistical Methods in Finance* (North-Holland, Amsterdam), 119-191.

- Hamilton, J. D., 1994, *Time Series Analysis* (Princeton University Press, Princeton, New Jersey).
- Harrison, M.J. and D. Kreps, 1979, Martingales and arbitrage in multiperiod securities markets, *Journal of Economic Theory* 20, 381-408.
- Harrison, M.J. and S.R. Pliska, S.R., 1981, Martingales and stochastic integrals in the theory of continuous trading, *Stochastic Processes and Their Applications* 11, 215-260.
- Harvey, A.C., E. Ruiz and N. Shephard, 1994, Multivariate stochastic variance models, *Review of Economic Studies*, 61, 247 – 264.
- Heston, S.L., 1993, A closed-form solution for options with stochastic volatility with applications to bond and currency options, *The Review of Financial Studies* 6, 327-43.
- Hol, E. And S.J. Koopman, 2002, Stock index volatility forecasting with high frequency data, Tinbergen Institute Discussion Paper, (Tinbergen Institute, Amsterdam, The Netherlands).
- Hull, J.C., 1997, *Options, Futures and Other Derivatives* (Prentice-Hall International).
- Hull, J.C. and A. White, 1987, The pricing of options on assets with stochastic volatilities, *Journal of Finance* 42, 2, 281-300.
- de Jong, F., 2000, Time series and cross-section information in affine term-structure models, *Journal of Business and Economic Statistics*, Vol. 18, No. 3, 300 – 314.
- Kim, S., Shephard, N. and S. Chib, 1998, Stochastic volatility: likelihood inference and comparison with ARCH models, *Review of Economic Studies* 65, 361-93.
- Lund, J., 1997, Econometric analysis of continuous-time arbitrage-free models of the term-structure of interest rates, Working paper, Aarhus School of Business (Aarhus, Denmark).
- Melino, A. and M. Turnbull, 1990, Pricing foreign currency options with stochastic volatility, *Journal of Econometrics* 45, 239-65.
- Mikosch, T., 1998, *Elementary Stochastic Calculus with Finance in View* (World Scientific Publishing).
- Sandmann, G. and S.J. Koopman, 1998, Estimation of stochastic volatility models via Monte Carlo maximum likelihood, *Journal of Econometrics* 87, 271-301.
- Scott, L.O., 1987, Option pricing when the variance changes randomly: theory, estimation and an application, *Journal of Financial and Quantitative Analysis* 22, 419-38.
- Shephard, N., 1996, Statistical aspects of ARCH and stochastic volatility, in: *Time series Models in Econometrics, Finance and Other Fields* (London, Chapman & Hall), 1-67.
- Stein, E.M. and J.C. Stein, 1991, Stock price distributions with stochastic volatility: an analytic approach, *The Review of Financial Studies* 4, 727-52.
- Taylor, S.J., 1994, Modeling stochastic volatility: a review and comparative study, *Mathematical Finance* 4, 183-204.
- Van der Ploeg, Antoine P.C., 2002, Volatility and derivative pricing in financial markets, master thesis, Tinbergen Institute Amsterdam and University of Amsterdam (Amsterdam, The Netherlands).
- Van der Sluis, P.J., 1999, Estimation and inference with the efficient method of moments with applications to stochastic volatility models and option pricing, Ph.D. Thesis, Tinbergen Institute (Amsterdam, The Netherlands).
- White, H., 1982, Maximum likelihood estimation of misspecified models, *Econometrica* 50, 1-25.
- Wiggins, J.B., 1987, Option values under stochastic volatility, *Journal of Financial Economics* 19, 351-72.